Universidade de Lisboa
Faculdade de Ciências da Universidade de Lisboa
Departamento de Biologia Vegetal

# Dissecting lineage priming in mouse embryonic stem cells

**Andreia Sofia Faria Pereira**

Trabalho orientado por:
**Doutor Domingos Henrique**
(Instituto de Medicina Molecular) e
**Professor Doutor Pedro Silva**
(Faculdade de Ciências da Universidade de Lisboa)

Mestrado em Biologia Molecular e Genética

Dissertação
2015

Universidade de Lisboa
Faculdade de Ciências da Universidade de Lisboa
Departamento de Biologia Vegetal

# Dissecting lineage priming in mouse embryonic stem cells

**Andreia Sofia Faria Pereira**

Trabalho orientado por:
**Doutor Domingos Henrique**
(Instituto de Medicina Molecular) e
**Professor Doutor Pedro Silva**
(Faculdade de Ciências da Universidade de Lisboa)

Mestrado em Biologia Molecular e Genética

Dissertação
2015

# Agradecimentos

Para mim, este último ano representou um conjunto de aprendizagens, a todos os níveis, que me fizeram crescer e serão úteis para o resto da minha vida, profissional e pessoal. E, por isso, deixo aqui o meu agradecimento a todos aqueles (sem exceção!) que tornaram isso possível.

Ao meu orientador externo, Domingos Henrique, que me deu a oportunidade de desenvolver a minha dissertação no seu laboratório e por me ensinar que quando o sábio está a apontar para as estrelas é nelas que nos devemos focar e não na mão do sábio;

Ao professor Pedro Silva, pela sua disponibilidade em coordenar este projecto ao nível da FCUL;

Aos meus colegas de laboratório, Alexandra, Elsa, Cláudia, Pedro e Sara, pelo acolhimento, apoio e disponibilidade que sempre demonstraram;

Um agradecimento muito especial à Ana Guedes, pessoa fundamental no desenvolvimento deste projecto. Sou profundamente grata, não só pelo ensino das técnicas e pela discussão dos resultados, mas também pela preocupação, disponibiliidade e amizade que demonstrou ao longo deste ano;

Aos meus amigos e colegas de mestrado, Inês, Pedro, Tiago e Vânia pelos momentos bons (e menos bons) que passamos juntos, mas, principalmente, pela amizade incondicional que demonstraram nestes dois anos;

À A. Francisca, A. Paula, Débora, Diana P., Inês S., Inês V., Joana, Laura, Marina, Patrícia, Tânia Daniela, Valentina pelo carinho e companheirismo ao longo destes anos; um especial agradecimento à Maria Teresa, amiga de todas e quaisquer horas, obrigada por toda a partilha de “xaine”;

Aos meus tios e primos, Carlos, Luísa, Rafael e Rúben, e à Filipa, Goreti, João, Luis, Marina, Paulo e Vitinho, por todo o carinho e alegria nos fins de semana deste último ano;

Ao Gonçalo, pelo carinho, ternura e paciência inesgotáveis;

E, por fim, a toda a minha família (sem exceção), pelos valores que me transmitiram; em especial aos meus pais e irmã, por serem para mim um exemplo de amor, honestidade, generosidade, coragem e perseverança. Obrigada por tudo!

## Abstract

Embryonic stem (ES) cells derived from the epiblast of mouse blastocysts are characterized by the ability to self-renew and to give rise to all embryonic lineages (a feature known as pluripotency). Therefore, mouse ES cells represent a unique tool to understand how transcription factor networks are modified during embryogenesis. Several knowledge has been accumulated concerning the gene regulatory networks and their transcriptional dynamics during pluripotency and lineage commitment. However, the understanding, at the single cell level, of how pluripotent transcriptional networks are dismantled, while lineage-affiliated genes networks emerge during lineage commitment of ES cells is still scarce.

The general aim of this study was to investigate single cell gene expression during neural commitment of pluripotent (*Oct4*, *Sox2*, *Nanog*), stochastic (*Car2*, *Cldn6*) and lineage-affiliated genes (*Cdh2*, *Sox3*, *Crabp2*, *Fgf5*, *Dnmt3b*, *T*). For that, I cultured mouse ES cells in conditions that allowed neural differentiation and performed single molecule RNA FISH analysis from cells collected at different time points of the differentiation protocol.

Here, I confirm that the highest expression of pluripotency genes occurs when mouse ES cells are cultured in pluripotent conditions. I also show that the pluripotency network is dismantled when cells enter neural lineage differentiation, due to decreased expression of *Oct4* and *Nanog*. In the first days of neural differentiation, neural genes start to be upregulated but their transcription is mainly bursty, similar to genes that are stochastically expressed in the first days of differentiation. Moreover, postimplantation epiblast markers (*Fgf5* and *Dnmt3b*) were found upregulated at day 3, when *Nanog* is also transiently upregulated, suggesting that a first transition from pluripotent mouse ES cells to a primed state similar to mouse epiblast stem cells occurs during early neural commitment. Later in neural differentiation, neural genes start to be expressed in a constitutive pattern and no expression of stochastic genes is observed, suggesting a second transition to irreversible neural commitment with formation of neural progenitors.

Overall, this work reveals molecular transitions and formation of intermediate subpopulations with specific gene signatures when mES cells are driven to neural fate.



Keywords: Embryonic Stem Cells, Pluripotency, Neural Differentiation, Single cell gene expression, Stochasticity

# Sumário

As células estaminais embrionárias de ratinho representam hoje em dia uma ferramenta extremamente últil para o estudo do desenvolvimento embrionário, uma vez que são capazes de se multiplicarem indefinidamente e são pluripotentes, isto é, são capazes de dar origem a todas as células das linhagens embrionárias. Estas células foram descritas como sendo semelhantes às células do epiblasto que compõem o blastocisto, antes da implantação uterina, por volta do dia embrionário (E) 4.5. Durante a implantação ocorrem várias mudanças morfológicas e de expressão genética no blastocisto e, por volta do dia E6.5, ocorre a gastrulação, processo onde as células do epiblasto pós-implantatório originam percursores das três linhagens embrionárias: mesoderme, endoderme e neuroectoderme.

A manutenção da capacidade de auto-renovação e pluripotência das células estaminais embrionárias é regulada principalmente por três fatores de transcrição: NANOG, OCT4 e SOX2. Estes fatores compõem o centro de uma vasta rede interativa de factores reguladores da pluripotência. Inicialmente, as células estaminais embrionárias eram vistas como uma população clonal de células com uma expressão genética homogénea destes fatores. No entanto, vários estudos vieram contrariar esta ideia demonstrando a sua expressão heterogénea nas células estaminais embrionária, bem como o impacto funcional desta heterogeneidade. Por exemplo, no caso de *Nanog*, as células estaminais que apresentam baixos níveis de NANOG têm tendência para expressar genes associados a linhagens embrionárias e, assim, estão mais propensas à diferenciação. Este processo, denominado de “lineage priming”, carateriza-se pela expressão esporádica e reversível de genes envolvidos na diferenciação em células pluripotentes e é considerado indispensável para a capacidade pluripotente das células estaminais, uma vez que confere plasticidade e competência para se diferenciarem.

Ultimamente, vários laboratórios têm-se dedicado a desvendar os possíveis mecanismos que desencadeiam esta heterogeneidade na expressão génica, sendo que flutuações estocásticas a nível da transcrição tem sido apontada como a principal causa. Um gene pode ser constitutivamente expresso, estando o seu promotor continuamente ativo ou o promotor pode alternar entre um estado ativo e inativo que desencadeia variações no número de trancritos desse gene ao longo do tempo. Esta alternância está presente nas células estaminais embrionárias, desencadeando variabilidade individual entre células que permite a distinção de subpopulações com diferentes níveis de susceptibiliadde para se diferenciarem.

Aquando da gastrulação, ocorre a formação do nó primitivo na parte posterior do embrião. A partir do nó primitivo, surge a linha primitiva de onde surgem os percursores da mesoderme e endoderme. Na parte anterior do embrião forma-se a placa neural anterior que dará origem à maior parte das células que compõem o sistema nervoso central. Quanto à

formação da placa neural posterior, que originará parte do sistema nervoso central, existem dois modelos que tentam responder a essa questão: o primeiro proposto por Niewkoop, em 1952, defende que a placa neural posterior deriva da regionalização de uma parte da placa neural anterior por sinalizadores de "posteriorização". Recentemente, outro modelo propõe a existência de uma população de progenitores neuromesodermais (PNMs) bipotentes na parte posterior do embrião, independente da placa neural anterior, que contribui para a formação da espinal medula, bem como da mesoderme paraxial, e, por isso, é caraterizada pela coexpressão de: *T*, marcador mesoendodermal e *Sox2*, marcador de células neuroprogenitoras.

Devido à difícil acessibilidade para estudar a gastrulação em embriões de ratinho, o desenvolvimento de protocolos de diferenciação *in vitro* têm sido cruciais para o estudo do surgimento das várias linhagens durante o desenvolvimento embrionário. Um desses protocolos consiste na diferenciação neural de células estaminais embrionárias aderentes em monocamada originando progenitores neuroepiteliais que se organizam em forma de rosetas, semelhante ao que acontece aquando da formação do tubo neural *in vivo.* A análise global da expressão genética ao longo deste protocolo já foi efetuada, permintindo a identificação de diferentes estadios que ocorrem quando as células estaminais embrionárias são direcionadas para a diferenciação neural. No entanto, é necessário um conhecimento mais detalhado da expressão genética a nível de células individuais.

Assim, o principal objetivo deste trabalho foi a análise da dinâmica transcricional de um conjunto de genes em células estaminais embrionárias individuais sujeitas a diferenciação neural. Os genes analisados pertencem a três classes distintas: genes pluripotentes (*Oct4*, *Nanog* e *Sox2*); genes associados a linhagens embrionárias, nomeadamente, genes neurais (*Cdh2*, *Sox3* e *Crabp2*), genes expressos no epiblasto pós-implantatório (*Fgf5* e *Dnmt3b*) e gene associado ao desenvolvimento mesoendodermal e marcador de PNM's (*T*) e genes estocásticos (*Car2* e *Cldn6*), isto é, genes sem uma função direta no desenvolvimento embrionário, mas com maior expressão em células com baixos níveis de NANOG comparando com os níveis em células com altos níveis de NANOG, em condições de pluripotência.

Para tal, 3 linhas de células estaminais embrionárias foram, inicialmente, mantidas em condições de pluripotência (BMP4 LIF) e, posteriormente, transferidas para um meio que promove a diferenciação neural, RHB-A. Em diferentes dias do protocolo de diferenciação, algumas células foram recolhidas e fixadas para single molecule RNA FISH (smFISH), um método que permite a quantificação de transcritos de um dado gene em células individuais.

Assim, foi possível observar, no estado de pluripotência, a existência de duas subpopulações: uma com elevada expressão de transcritos dos 3 fatores de transcrição de pluripotência, e outra com níveis intermédios de transcritos de *Oct4* e baixos níveis de transcritos de *Nanog* e *Sox2* . Quando as células são direcionadas para a linhagem neural,

ocorre a desestruturação da rede de fatores de transcrição reguladores da pluripotência devido à diminuição da transcrição de *Oct4* e *Nanog*. No entanto, ao dia 3 do protocolo de diferenciação neural, 10% das células ainda expressam níveis elevados de transcritos de *Nanog* e *Oct4*, representando uma população de células ainda pluripotentes. O número de transcritos de *Sox2* por célula sofre apenas um ligeiro decréscimo, mantendo-se significativamente expresso durante a diferenciação neural, o que confirma a importância deste gene quer na manutenção da pluripotência, quer no desenvolvimento neural.

Quanto à expressão de genes estocáticos (*Car2* e *Cldn6*) ao longo do protocolo de diferenciação neural, esta é estocástica, havendo uma maior frequência de células que expressam altos níveis de transcritos destes genes ao dia 1; no entanto, ao dia 6 do protocolo de diferenciação, a frequência de células que expressam altos níveis destes genes é praticamente nula, sugerindo que a transcrição destes genes é restrita aos primeiros dias de diferenciação. No estado de pluripotência, as células com maior expressão de *Car2* e *Cldn6* apresentam simultaneamente níveis baixos de transcritos de *Nanog*, sugerindo um estado de "priming" nestas células pluripotentes. Foi ainda possível verificar que, ao dia 1 da diferenciação neural, ao contrário dos dias 0 e 3, a elevada expressão destes genes era independente, o que sugere uma janela temporal onde as células são mais permissivas a explorar o genoma, resultado da configuração "aberta" da cromatina e do decréscimo da expressão dos factores de transcrição que regulam a pluripotência.

A análise da expressão dos genes neurais em células individuais revelou que, nos primeiros dias de diferenciação, *Cdh2* e *Sox3* apresentam uma transcrição estocástica. No entanto, em dias mais avançados do protocolo é possível extrapolar que a transcrição destes genes é feita de uma forma constitutiva. Esta mudança no padrão de expressão de *Cdh2* e *Sox3* sugere que a expressão estocástica destes genes nos primeiros dias de diferenciação é um potencial mecanismo de "priming", que ocorre antes das células se compromoterem para a linhagem neural. Estas observações confimam a hipótese que o compromisso irreversível para uma dada linhagem é dependente de dois eventos consecutivos: a) ativação da transcrição de genes associados a uma dada linhagem e b) expressão desses genes acima de um dado nível.

Quanto ao estudo da expressão em células individuais de marcadores de epiblasto pós-implantatório, *Fgf5* e *Dnmt3b*, é possível constatar um aumento na frequência de células que expressam altos níveis de transcritos destes genes, do dia 1 para o dia 3 do protocolo de diferenciação neural. Este aumento está associado à reexpressão transiente de *Nanog* ao dia 3, sugerindo o aparecimento de uma população de células com expressão genética semelhante a células do epiblasto pós-implantatório.

Em conjunto, estes resultados sugerem que os primeiros dias de diferenciação neural são caracterizados pela desintegração da rede de fatores reguladores da pluripotência, bem

como por um contínuo aumento da expressão de genes neurais e transiente expressão de genes estocásticos e característicos do epiblasto pós-implantatório. Este estadio precede o estadio em que as células estão irreversivelmente comprometidas para a diferenciação neural.

Por fim, quando avaliando a existência de PNM's, pela co-expressão de *Sox2* e *T*, o número de células com altos níveis de transcritos destes dois genes foi muito baixo, devido à transcrição residual de *T*. Esta observação pode dever-se ao facto de o meio neural RHB-A ser bastante selectivo, inibindo a sobrevivência de células que expressam genes de linhagens não-neurais. Além disso, este meio, leva à diferenciação de células com características de tecidos neurais anteriores, o que dificulta o surgimento de células com expressão semelhante aos PNM's.

Palavras-chave: Células Estaminais Embrionárias, Pluripotência, Diferenciação neural, Expressão génica em células individuais, Estocacidade

# Contents

# 1 Introduction

## 1.1 Early Mouse Embryonic Development

Like in other mammals, mouse embryogenesis starts with one single totipotent cell, the zygote, from which both embryonic and extraembryonic tissues are originated. Later, a sixteen cell structure is formed, the morula (Figure 1). During this stage, asymmetric cell divisions occur, and "outside" and "inside" cells are generated. These cells differ not only in their position within the embryo, but also in their molecular properties and fate potential. While "outside" cells will give rise to the extraembryonic trophoectoderm (TE), the precursor of placenta; "inside" cells will give rise to the inner cell mass (ICM) (reviewed in [1]). These two cellular structures compose the early preimplantation blastocyst, formed around embryonic day 3.5 (E3.5) (Figure 1). The next cell fate decision occurs in the ICM, with segregation of the cell population that will originate the embryo proper, the pluripotent epiblast, and the cell population that will mainly constitute the yolk sac, the hypoblast or primitive endoderm (PE) (Figure 1). The foundation for this segregation is still in debate but it has been suggested that stochastic fluctuations in gene expression, followed by a signal reinforcement are sufficient to drive this second lineage choice[2]. Another hypothesis states that the timing of cell internalization upon the derivation of the ICM (first cell fate decision) affects the subsequent allocation of epiblast and PE cells[3].

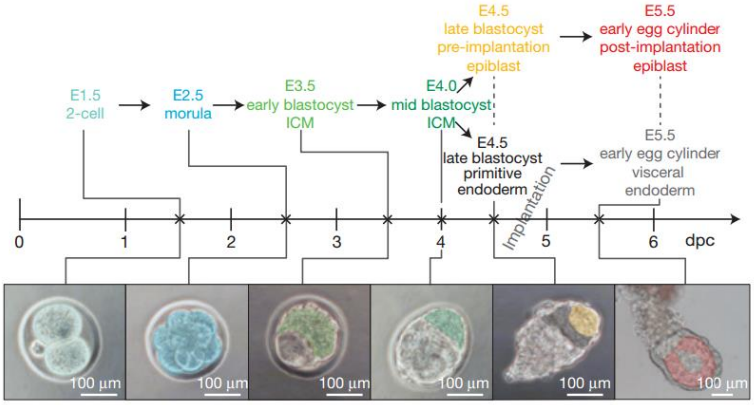



**Figure 1**-Scheme representative of the time line and morphology of the early mouse embryonic development. The cells marked with pseudocolors represent the developmental stage indicated with the same pseudocolors in the descriptions. dpc, days post coitum. Adapted from Boroviak et al.,2014 [7].

By the time of implantation (~E4.75), the blastocyst is organized into three distinct cell lineages: TE, PE and the epiblast. After implantation, the mouse blastocyst undergoes dramatic morphological alterations, losing its spherical shape and acquiring an elongated form, termed "egg cylinder" (Figure 1). The epiblast cells divide rapidly and reorganize in a pseudostratified epithelium located inside the egg cylinder (reviewed in [4]). Around E6.5, the epithelialized epiblast undergoes gastrulation, with generation of the three embryonic lineages (mesoderm, endoderm and ectoderm).

## 1.2 Mouse Embryonic Stem Cells

ES cells, either derived from human or mouse embryos, have two unique features: ability to self-renew and pluripotency, the capability to give rise to all cell types of the three embryonic lineages.

Isolation of mouse embryonic stem (mES) cells was firstly reported in 1981[5,6], and until recently the association between mES cells and the embryo was controversial. Recently, however, Boroviak *et al*[7] reported a close relationship between *in vitro* mES cells and *in* vivo E4.5 epiblast cells[7], based on global transcriptional and functional data.

Besides the derivation of ES cell lines, it is also possible to originate other pluripotent cell lines from different embryonic developmental stages: postimplantation epiblast stem (EpiS) and embryonic germ (EG) cell lines. ES and EG cells share some features including colony morphology and gene expression pattern. On the other hand, EpiS cells exhibit different colony morphology and a different gene expression profile, when compared to ES cells (reviewed in [8,9]). In order to discriminate these differences, mES and mEpiS cells are termed "naïve" and "primed" pluripotent cells, respectively, representing two distinct states of pluripotency. These cells exhibit common and divergent characteristics: both express pluripotent transcription factors OCT4 (octamer-binding transcription factor 4) and SOX2, and are able to originate cells of all embryonic lineages *in vitro* and *in vivo*. However, mEpiS cells are derived from a more developmental advanced epiblast (postimplantation, E5.5) than mES cells (preimplantation, E4.5). Therefore, it is believed that mEpiS cells are one step closer to lineage commitment than mES cells[10,11]. Following this notion, mES cells are characterized by expression of preimplantation blastocyst markers like *Nanog*, *Rex1*, *Stella*, *Tbx3*, *Dax1* and *CD3*. In contrast, these markers are downregulated in mEpiS cells, which show expression of postimplantation epiblast markers (*Fgf5*, *Otx2*, *Dnmt3b* and *Nodal*), as well as upregulation of lineage-affiliated genes (*T,* also known as *Brachury*, *Sox17*, *Eomes* and *Foxa2)* (reviewed in [8]). mEpiS cell lines are also different in chromatin configuration, displaying X chromosome inactivation, an epigenetic mark of differentiation, and global increase DNA methylation due to enhanced *Dnmt3b* gene expression, two features that are absent from mES cells (reviewed in [12]).

### *1.2.1 In Vitro Culture Systems of mES cells*

As an *in vitro* counterpart of the preimplantation epiblast, mES represent a unique tool to investigate embryonic development. Therefore, the study of signalling pathways that enable to capture and sustain the naïve state of mES is crucial in this field.

mES cells can be maintained *in vitro* using different culture media. When the first pluripotent cell lines were derived[5,6], cells were cocultured with mitotically inactive mouse fibroblasts, so called "feeder" cells because they provide signalling molecules, supplemented with fetal calf serum, which inhibits mES cells differentiation. Later, it was found that Leukemia Inhibitory Factor (LIF) was the cytokine secreted by "feeder" cells that enable mES cells self-

renewal[13,14]. LIF binds to the GP130 receptor and through activation of the transcription factor STAT3 drives long-term self-renewal of mES cells[15]. Nowadays, the standard culture medium for mES cells includes serum supplemented with LIF (Serum LIF conditions), but mES cells cultured in this system present some spontaneous differentiation, strongly influenced by serum batch and cell density (reviewed in[16]). In 2003, Ying et al[17] reported the maintenance of mES cells in serum-free conditions by adding Bone Morphogenetic Protein 4 (BMP4) and LIF to a chemically defined culture medium (BMP4 LIF conditions). BMP4 acts by activating Inhibitor of differentiation (Id) proteins, through the Smad pathway, leading to repression of neural lineage commitment, while LIF/STAT3 pathway inhibits mesoendodermal differentiation. More recently, in 2008, another stem cell culture system was reported, the "2i" media[18]. This is composed of 2 inhibitors, CHIRON99021 (CHIRON) and PD0325901 (PD03), in a chemically defined media. PD03 is a MEK inhibitor and blocks phosphorylation of ERK1/2, promoting long-term self-renewal of mES cells. However, PD03 is insufficient to support mES cells' viability in the absence of LIF. Therefore, addition of a GSK3 inhibitor, CHIRON, stabilizes mES cells' proliferation. Using these two inhibitors (2i conditions) with optional addition of LIF, it is possible to keep mES cells in a pluripotent ground state shielded from differentiation stimuli[11,18,19].

### *1.2.2 Pluripotent Gene Regulatory Network of mES cells*

Pluripotency is initiated and maintained due to interactions between transcription factors, chromatin modifiers and other regulatory genes. According to some studies, when OCT4[20], NANOG[21] and SOX2[22] are depleted in mouse embryos, epiblast formation is severely compromised and early differentiation is triggered, leading to the assumption that these transcription factors are critical regulators of pluripotency [19].

OCT4, also known as POU5F1, is a POU homeodomain transcription factor. It is expressed in the ovum and after fertilization is distributed to all blastomeres. After the first cell fate decision, *Oct4* expression becomes restricted to the ICM. In homozygous *Oct4*-deficient mouse embryos, no mature ICM is formed and the blastocyst develops into a mass of giant trophoblast cells [20]. Upon the second cell fate decision, *Oct4* is expressed in the epiblast and transiently in the PrE, but after implantation its expression is confined to epiblast cells [23]. *Oct4* expression is maintained after implantation but continuously decreases during gastrulation, becoming undetectable in the somatic cells of 12- to 15-somite embryos (E8.75), and therefore marking the loss of pluripotent capacity [24]. In mES cells, proper levels of *Oct4* expression are critical for the maintenance of its pluripotentiality, with an increased expression leading to PrE and mesoderm differentiation and a decrease in *Oct4* expression levels triggering loss of pluripotency and dedifferentiation to trophoectoderm [25].

NANOG (designation derived from the Irish mythological land "Tír na nÓg", or "Land of the Young" [26]) is a homeodomain transcription factor. Unlike OCT4, NANOG is not maternally

inherited, but appears in a random cellular distribution at the end of the morula stage [27]. Later, *Nanog* becomes exclusively expressed in the ICM, with a mosaic distribution. In the early ICM, *Nanog* and the early PrE marker *Gata6*, are co-expressed, although, upon segregation of epiblast and PrE cells, the expression of *Nanog* and *Gata6* becomes mutually exclusive, with NANOG expressing-ICM cells originating epiblast cells and GATA6 expressing-ICM cells forming PrE cells (reviewed in[1,28]). At the time of implantation, *Nanog* expression is downregulated and re-emerges in the posterior egg cylinder (reviewed in [19]). Upon gastrulation, *Nanog* expression declines, becoming undetectable in somatic cells of 3- to 5-somite embryos, earlier than *Oct4* [24]. However, *Nanog* and *Oct4* expression persist in primordial germ cells [29]. *In vitro*, *Nanog* expression was found to be important in mES cells self-renewal, with its upregulation being sufficient to maintain clonal propagation of pluripotent cells in the absence of LIF cytokine signalling [21,26]. *Nanog* expression is not fundamental to maintain mES cells in a pluripotent state but when downregulated turns mES cells more susceptible to differentiate [30,31].

*Sox2* encodes a transcription factor member of the Sox (SRY-related HMG box) family and belongs to the SoxB1 subgroup, which also includes *Sox1* and *Sox3*. Like OCT4, SOX2 is also maternally inherited with zygote-derived *Sox2* mRNA starting to be detected in some cells at morula stages (E2.5) and in the blastocyst. Its expression persists throughout the ICM and later in epiblast cells. Around E7.5, most of its expression becomes restricted to the anterior part of the embryo that will develop into the anterior neuroectoderm [22,24]. Its expression was also found in a posterior region, when posterior bipotent neuromesodermal progenitors (NMp's) are established (reviewed in [32]). Like OCT4 and NANOG, upon somitogenesis, *Sox2* expression persists only in primordial germ cells [22,24]. In mES cells, SOX2 was described as a stabilizer of the pluripotent state through regulation of *Oct4* expression levels [33]. Downregulation of *Sox2* expression promotes the differentiation of mES cells into trophectoderm-like cells[34], whereas its overexpression predisposes mES cells to neuroectodermal differentiation [35].

These three transcription factors, OCT4, SOX2 and NANOG, are known as "core" pluripotency factors in human and mouse embryonic stem cells, and several studies point for direct and indirect interactions between them and other transcription factors, establishing a pluripotent regulatory network[34,36,37]. Several studies indicate the existence of other factors also involved in pluripotency regulation, such as ESRRB, STAT3, TBX3 and KLF4 (reviewed in [38]).

### *1.2.3 Heterogeneous gene expression in mES cells*

Initially, ES cultures were seen as homogenous population of cells, with equal differentiation potential. However, further evidence suggested that mES cell cultures comprise distinct cell types, in terms of differentiation potential, chromatin configuration and gene expression. In fact, *Nanog* was found to be heterogeneously expressed in individual mES cells [31,39,40] and *Oct4*, firstly described as homogenously expressed, also vary among individual mES

cells [41]. However, some authors reply that mouse ES cells heterogeneity may be a culture epiphenomena, as a consequence of a disordered signalling environment created by *in vitro* conditions[42]. In fact, when mES cells expressing an unstable fluorescent reporter for NANOG are cultured in Serum LIF conditions, only 50% present NANOG expression; on other side, when mES cells are transferred to a 2i culture medium, a significantly more homogenous population of cells is generated, with 90% of mES cells expressing Nanog:VNP [31]. However, fluctuations in *Nanog* expression still occur in 2i media [31]. Opposing this view, heterogeneity was shown to have functional impact in mES cells differentiation potential. Abranches et al. [31] showed that Low-NANOG mES cells are more susceptible to differentiation and markedly express lineage-affiliated genes when compared to High-NANOG mES cells. This higher expression of lineage-affiliated genes in Low-NANOG cells has been associated with "lineage priming", a process where pluripotent cells display sporadic and reversible expression of these genes, reflecting an increased predisposition to enter differentiation. "Lineage priming" is considered indispensable for the pluripotency capacity of mES cells, conferring the developmental competence and plasticity to subsequently differentiate. Several studies have reported that pluripotent cells *in vitro* oscillate between a ground state and another where cells are more prone to lineage commitment [44,45].

Not only *Oct4* and *Nanog* have been proven to be heterogeneously expressed in mES cells, but also other pluripotency transcription factors like *Stella*, *Rex1*, *Esrrb*, *Klf4* and *Tbx3* (reviewed in [8,40]). Interestingly, this mosaic expression is not only observed for pluripotency transcription factors but also for lineage-affiliated genes such as *T*, mesoendodermal marker, and *Hhex*, hypoblast marker (reviewed in [8]).

But what drives this heterogeneous gene expression in mES cells?

## 1.3 Stochastic and noisy gene expression in mouse ES cells

There are a number of possible mechanisms to explain heterogeneities in gene expression. One of the sources is the unequal provision of extrinsic signals to mES cells that could generate correspondingly heterogeneous pluripotency factor expression; or even autoregulatory feedback loops of pluripotent genes also regulate heterogeneity. Nevertheless, stochastic fluctuations in mRNA levels in single cells is often viewed as the major source of heterogeneity of gene expression in mES cells.

Generally two models of transcription could be distinguished (Figure 2): fully activated genes that display constant transcription and therefore the number of mRNA molecules produced over time is constant (Figure 2a, 2c); and those that are only activated in certain contexts with the number of mRNA molecules varying along time (Figure 2b, 2d). The later model of transcription is usually characterized by fluctuations in gene expression, triggering a pulsatile transcription. If transcription occurs with bursts, variations in the number of mRNA

molecules present in each cell is expected. These bursts start and end when a gene transits from an inactive to active state and vice-versa, respectively, and these transitions were proven to be due to intrinsic noise[46]. According to Elowitz et al[47], there are two types of noise in gene expression: extrinsic and intrinsic. Extrinsic fluctuations refer to variations in the number of RNA polymerases, ribosomes, transcriptional activators and repressors and affect the transcription of the entire genome; intrinsic fluctuations are due to the randomness inherent to transcription and translation. Both sources of noise are significant. However, intrinsic noise is the major cause of transitions between active/inactive state. Furthermore, the activation of a certain gene causes burst-like transcription in all genes closer to that genomic locus, whereas genes located in distant genomic regions burst independently [46]. Like for other mammalian cells, transcriptional noise is a hallmark of mES cells and is essential for generation of heterogeneity in cell populations. In mES cells, as presented in section 1.2.3, cell-to-cell variations resulting from transcriptional noise are important to keep a subpopulation of pluripotent mES cells continuously primed to differentiation, providing the opportunity for cells respond to several external signals within a short period of time.

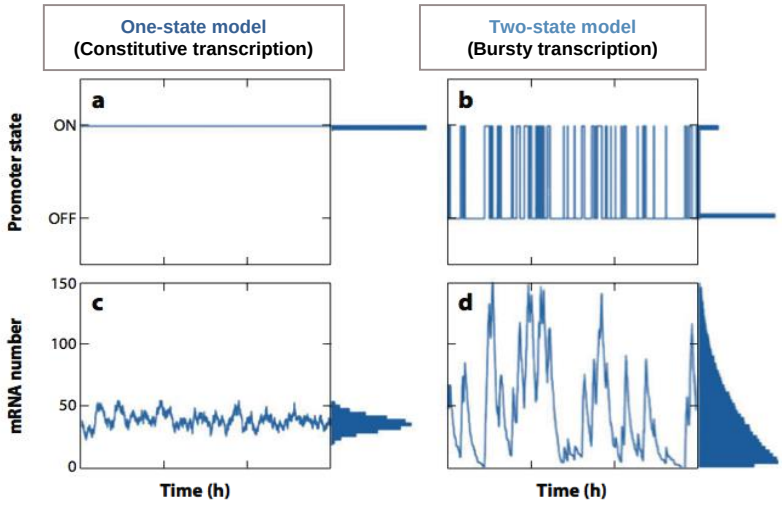



**Figure 2-** Models for gene transcription. One-state model applies to genes expressed constitutively and therefore always with an "ON" promoter state (a) and constant number of mRNA molecules produced overtime resembling a Gaussian distribution(c). Two-state model represents genes with a bursty transcription where the promoter state oscilates between "ON" and "OFF"(b) and therefore the number of transcripts produced overtime is not constant approaching a long tailed distribution(d). Adapted from Raj and van Oudenaarden, 2008 [55].

## 1.4 From Pluripotency to Differentiation

When mES cells commit to differentiation, bursty transcription is still present [48] and therefore heterogeneity of increasingly expressed lineage-affiliated genes is observed (reviewed in [8,40]). The same is valid for genes that do not present any developmental role, but are highly expressed in Low-Nanog comparing to High-Nanog mES cells (RNA-seq data; E. Abranches, unpublished data). In this work, these genes are designated as "stochastic" genes, to distinguish them from lineage-affiliated genes.

In the mouse embryo after implantation, as described in section 1.1, the pluripotent epiblast undergoes morphological and transcriptional modifications. Around E6.5 (Figure 1), gastrulation begins with the formation of the node that acts as a body plan organizer, on the posterior side of the epiblast. From the node, a structure called primitive streak is formed. With the formation of the primitive streak, the anterior-posterior axis becomes morphologically obvious with the streak located on the posterior side of the embryo. The cells that move through the streak become mesendoderm, which are the precursors of mesoderm and endoderm cells. Anteriorly, the acquisition of neural fate is achieved by the induction of the anterior neural plate derived from the anterior epiblast, while the posterior neural plate arises from regionalization of the anterior neuroectoderm through "posteriorizing" signals. This model is derived from the "activation-transformation" hypothesis proposed by Nieuwkoop in 1952 [49]. More recently, the existence of a population of bipotent neuromesodermal progenitors (NMp's) was reported in the posterior epiblast in E8.5 mouse embryos [32]. These cells contribute to both spinal cord and paraxial mesoderm, and are characterized by the co-expression of early mesodermal marker *T* and neural progenitor marker SOX2 (reviewed in [32]). Therefore, an alternative model[32] proposes that neural induction occurs through formation of the anterior neural plate, while some posterior neural tissue is derived independently from the anterior neural plate. Several studies have reported the derivation of *in vitro* NMp-like cells, providing a unique tool to study the formation, gene regulatory networks and developmental potentialities of these cells.

### *1.4.1 Neural differentiation of mES cells*

*In vitro* differentiation of mES cells cultured in defined conditions provides a tractable system to dissect and understand the process of exit from naive pluripotency and entry into lineage specification. ES cells can be driven into neural differentiation by the formation of multicellular aggregates, embryoid bodies (EBs), or by adherent monolayer cultures[51,52]. In the later system, mES cells are cultured in feeder-free conditions in the presence of serum-free medium, to maintain cells in an undifferentiated state. When these cells are exposed to a defined medium, RHB-A, that drives mES cells to an anterior neural fate, neural progenitors (NPs) are generated, which become organized in rosette-like structures[52] (Figure 3). These structures resemble the formation of the neural tube with a clear apical-basal polarity marked by the apical expression of junction proteins like N-Cadherin and ZO-1[52,53].

During this *in vitro* neural differentiation protocol, a global characterization of the gene expression signatures present in each particular stage of neural development was performed, allowing the identification of the main stages and transitions that occur when mES cells are driven to neural fate. However, a deeper knowledge on how gene expression in individual cells was affected during neural commitment and differentiation was lacking.

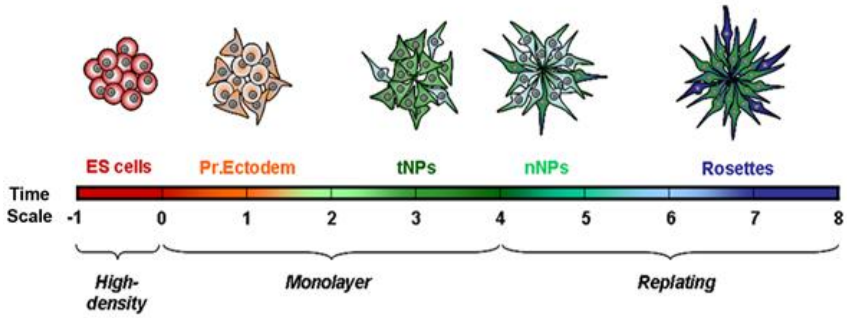



**Figure 3**- Diagram representative of the successive cellular states that occur along the monolayer neural differentiation based on their transcriptional profiles. Initially mES cells are found in the pluripotent state (Days -1 and 0); when plated in RHB-A medium, in the first days cells acquire gene expression similarities with Primitve Ectoderm cells (Pr. Ectoderm). Then, a population of transient NPs (tNPs) emerges and latter give rise to permanent neural NPs (nNPs). From day 6 onwards, this population will originate the set of NPs organized in rosette-like structures (Rosettes). Adapted from Abranches et al. [52]

Recently, in 2013, Trott and collaborators [54], reported the study expression of certain genes when mES cells exit pluripotency, being cultured in N2B27, a neural basal medium, at the single cell level. In this study, it was reported a decrease in *Rex1*, *Nanog* and *Oct4* expression, as well as transient expression of the epiblast markers *Fgf5* and *Otx2* and increase in the expression of neural genes was also observed [54].

Nevertheless, the number of analysed cells and the studied gene expression correlations in this report were scarce, not allowing the distinction of putative different subpopulations that could arise within the same stage of neural differentiation. Furthermore, the transcription mode (constitutive or bursty) of the genes analysed was not examined and so, the transcriptional dynamics of genes with different developmental roles, in single cells, during *in vitro* neural differentiation, is still an unresolved question.

# 2 Aims

The main aim of this project was to analyse the single cell transcriptional dynamics of selected genes (pluripotent, lineage-affiliated and stochastic genes) when mES cells are committed to neural differentiation. The specific aims were:

- Quantify, at the single cell level, the expression of mRNA molecules from a set of selected genes (pluripotent, lineage-affiliated and stochastic genes) along monolayer neural differentiation, and apply correlation analysis for the distributions obtained;
- Understand how is the core pluripotency network dismantled along neural differentiation;
- Understand how transcription dynamics of lineage-affiliated and stochastic genes changes along neural differentiation;
- Assess the presence of cells co-expressing *T* and *Sox2* (NMp's markers) along neural differentiation.

# 3 Material and Methods

## 3.1 Materials

### *3.1.1 Reagents*

The reagents, solutions/media, antibodies and single molecule RNA FISH probes used in this study are listed in Table S1, Table S2, Table S3 and Table S4, respectively.

### *3.1.2 Embryonic stem cell lines*

The three mouse ES cell lines used in the experiments were: *E14Tg2a*, *Nd*[43] (Nanog dynamics, derived from *E14Tg2a*, with a short lived VNP reporter under the control of *Nanog* regulatory regions) and *46C* (derived from *E14Tg2a*, with a GFP reporter under the control of *Sox1* promoter region). *E14* (for *E14Tg2a*) and *46C* were a gift from Meng Li-MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College, London, UK- and Austin Smith-Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Cambridge, UK).

## 3.2 Methods

All of the experiments described in sections 3.2.1 and 3.2.2 were performed in a sterile laminar flow hood class II, type A/B3.

### *3.2.1 Expansion of Embryonic stem cells*

For mES cell expansion, cells were cultured on gelatin coated dishes (10 minutes coating), in GMEM (Table S2) supplemented with 10% FBS and LIF (2ng/mL) (Serum LIF conditions), in a 5% $CO_2$ incubator. The frozen cell stock (liquid nitrogen) was thawed in pre-heated supplemented GMEM and centrifuged at 1200rpm for 4 minutes. After pellet re-suspension in supplemented GMEM, cells were plated on 0.1% (v/v) gelatin-coated dishes. Medium was changed 6 hours later to eliminate DMSO residues. The cells' morphology was assessed daily by direct visualization on a bright field microscope, and cells were passaged every other day, at a constant plating density of $3x10^4$ cells/$cm^2$. For each passage, cells were washed twice with PBS and dissociated with 0,1% trypsin for 2min at 37ºC. Cells were immediately re-suspended in supplemented GMEM in order to neutralize trypsin, centrifuged at 1200 rpm for 4min and re-suspended again in supplemented GMEM. In each passage, cells were counted using trypan blue dye exclusion method and the required amount of cells was re-plated as described before. At each passage Nanog:VNP and Sox1-GFP expression was assessed by Fluorescence Activated Cell Sorting (FACS) (section 3.2.1.2).

To prepare stocks of mES cells to be frozen, $3x10^6$ cells were frozen in supplemented GMEM with 10% Dimethyl sulphoxide (DMSO) and stored in liquid nitrogen. Every time cells were frozen, a sample was collected to test for Mycoplasma contamination (section 3.2.1.1).

#### *3.2.1.1 Polymerase Chain Reaction (PCR) and Electrophoresis for Mycoplasma detection*

Each time a new mES cell stock was prepared, the presence of mycoplasma contamination was determined by PCR. For that $1.0x10^6$ cells were collected, centrifuged at 2000 rpm for 5min, re-suspended in wash buffer (Table S2) and centrifuged again in the same conditions. Then a 1:1 mix of solution A and solution B (Table S2) was added to the cells and the mixture was incubated for 1h at 60ºC followed by 1 hour at 90ºC.

PCR was performed using rTaqPolymerase, amplifying a conserved region in the 16S RNA *Mycoplasma* ribosomal gene. The amplification was performed with an initial step of denaturation at 95ºC for 5 min, followed by 30 cycles of denaturation at 95ºC for 30s, annealing at 58ºC for 1.5 min and extension at 72ºC for 1.5 min, and a final step of extension at 72ºC for 10 min. The reactions were prepared for a final volume of 25 µL: 3µL of sample, 1x buffer, 0.2mM dCTP, 0.2mM dGTP, 0.2mM dATP, 0.2mM dTTP, 25pmol of each primer (Table 1) and 2.5U of rTaq Polymerase.

After PCR reaction, the PCR products were analysed in a 1.5% agarose gel. Gels were prepared by heating agarose in 1x TAE buffer until complete dissolution, followed by addition of gelRed (1:20 dilution) to allow visualization of DNA fragments. PCR products were mixed with loading buffer in a 5:1 proportion and loaded on an Agarose gel. An electric voltage of 70 volts was applied to the gel for 90min and DNA fragments were visualized under ultraviolet light at 260nm or 365nm, using Bio-Rad Image Lab Software.

At the same time, the quality of the DNA preparation was confirmed by performing a PCR to detect GADPH gene, a housekeeping gene, used as an internal control. Moreover, a plasmid carrying the *Mycoplasma* 16S ribosomal RNA gene was used as a positive control and an ultrapure water sample as a negative control. The size of the fragments was estimated by comparison with linear DNA strands of known molecular weight (1kb Plus DNA Ladder-Invitrogen).

**Table 1**: Oligonucleotides used for mycoplasma detection.

| Gene | Sense primer | Antisense primer | Product size (bp) | Annealing temperature (ºC) | Company |
|---|---|---|---|---|---|
| ***Mycoplasma*** | TGCACCATCTGTCACTC TGTTAACCTC | ACTCCTACGGGAGG CAGCAGTA | 717 | 58 | Sigma |
| ***GAPDH*** | ATTCAACGGCACAGTCA AGG | TGGATGCAGGGATG ATGTTC | 580 | 60 | Sigma |

#### *3.2.1.2 Fluorescence Activated Cell Sorting (FACS)*

In order to quantify the percentage of cells that express VNP and GFP reporter in the various cell culture conditions here used, fluorescence activated cell sorting (FACS) was

performed. Cells were dissociated and $5x10^5$ cells were re-suspended in PBS. Live cells were gated based on forward and side scatter and by propidium iodide dye exclusion. Data was acquired in a FACS Calibur and in each data acquisition, 10000 gated events were recorded and the data obtained was subsequently analysed using *FlowJo* software.

### *3.2.2 Neural differentiation in adherent monolayers cultures*

Neural differentiation of mES cells was done according to a protocol described by Abranches et al[52]. Briefly, cells were plated in 0.1% (v/v) gelatin-coated dishes at high density ($1x10^5$cells/$cm^2$) (Day -1) in serum-free medium (ESGRO clonal grade) supplemented with LIF (2ng/mL) (BMP4 LIF conditions). After 24hours (Day 0), morphology and cell confluence were evaluated in a bright field microscope. Cells were dissociated, counted and plated in gelatin-coated dishes in a neural differentiation medium (RHB-A) at density of $1.5x10^4$cells/$cm^2$. For each day of the differentiation protocol, cell morphology was observed and documented. Medium was renewed at day 2. At day 4, cells were dissociated, counted and re-plated in PDL-Laminin coated dishes (section 3.2.2.1), at a density of $2x10^4$ cells/$cm^2$ in RHB-A supplemented with 5ng/mL murine bFGF, and medium was changed again at day 6. At each timepoint of the neural differentiation protocol, Nanog:VNP and Sox1-GFP reporters expression was assessed by FACS (section 3.2.1.2).

#### *3.2.2.1 PDL-Laminin coating*

The tissue culture dishes were covered with PDL solution (10µg/mL in PBS) and left for 1h at room temperature. The dishes were then washed twice with PBS and covered with Laminin solution (2.5µg/mL in PBS) and left overnight in the 5% $CO_2$ incubator at 37ºC.

According to the analysis to be performed, coating was either done directly on culture dishes (for cell fixation for smFISH purposes - see section 3.2.2.3 - or for protein analysis using FACS - see section 3.2.1.2) or on glass coverslips (for immunofluorescence analysis purposes - see section 3.2.2.2).

#### *3.2.2.2 Immunofluorescence*

Immunofluorescence was performed to visualize the presence and structural organization of rosettes-like structures, confirming the presence of neural progenitors at day 6 of monolayer neural differentiation protocol.

Cells grown on coverslips were washed twice with PBS, for 5min, and fixed with 4%(w/v) paraformaldehyde (PFA) in PBS, for 15min at 4ºC. Cells were washed twice for 5min and residual PFA was washed with fresh 0.1M Glycine in PBS for 10min at room temperature (RT). Cells were then permeabilized by incubation with 0.1% Triton in PBS for 10min at RT and blocked with blocking solution (Table S2) for 30min at RT. Primary antibodies (Table S3) were

diluted in blocking solution and incubated overnight (O/N) at 4ºC. In the next day, cells were washed three times with TBST for 5min each. Appropriate secondary antibodies (Table S3) were diluted in blocking solution and incubated for 30min at RT. Cells were washed three times for 5min with TBST and counterstained with 4',6-diamidino-2-phenylindole (DAPI) for 5min. After washing three times in PBS for 5min, cells in coverslips were mounted with Mowiol mounting medium.

Images were acquired using a Leica DM5000B widefield fluorescence microscope with a monochrome CCD camera and image analysis was performed using Image J and Adobe Photoshop software.

#### *3.2.2.3 Single Molecule RNA Fluorescent In Situ Hybridization (smFISH)*

For smFISH, cells were dissociated, washed with PBS and fixed in 4% formaldehyde for 10min at RT. After that, cells were washed twice in PBS in re-suspended in 2 to 4mL of Ethanol 70%. These fixed cells can be kept at 4ºC for long periods of time.

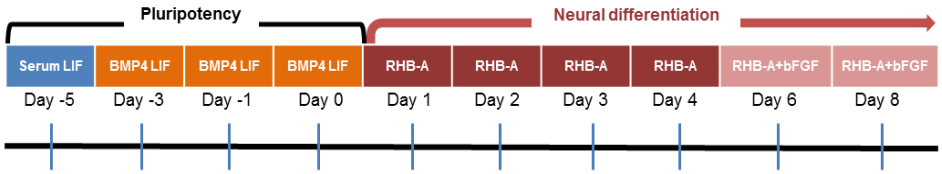
For each smFISH experiment, 100-200µL of fixed cells were re-suspend in 1mL of wash buffer (see Table S2). After that, cells were incubated O/N at 37ºC in a mix composed of 100µL hybrization buffer (Table S2) and 1 µL of each Stellaris TM FISH probe (Table S4)[55]. Up to three different probes, targeting 3 different mRNA transcripts were used in a single smFISH experiment. Probes were previously dissolved in TE (Table S2) to create a probe stock at a global concentration of 1 to 12µM. Each probe set is composed of 25 to 45 unique oligonucleotides (20 nucleotides each) complementary to a different region of the target RNA. Each probe is labelled with a fluorophore: Alexa594, Cy5 or Tmr, with excitation peak/ emission wavelength of 590/ 617nm, 649/ 670nm and 564/ 570nm, respectively. Each probe set produces enough signal to detect one single mRNA molecule as one spot during image acquisition.

In the next day, cells were washed twice: firstly with wash buffer for 30min at 37ºC and then with wash buffer and DAPI (1µL of 1mg/mL solution) for 30min at 37ºC. Next, cells were washed in glox buffer solution and then finally re-suspended in an antifade solution (Table S2). After suspension cells were mounted between slide and coverslip and carefully smashed to decrease the volume of the cells and allow increase of signal quality and finally properly sealed.

Cells were imaged within 24hours on an inverted fluorescence Zeiss Axiovert 200M microscope, using a 100x1.4 oil-immersion objective, a cooled CCD camera (Roper Scientific Coolsnap HQ CCD) and filtersets suitable to the excitation and emission wavelengths of the applied fluorophores. For each image 30 Z stacks, with 0.3µm step size were recorded. Data were processed using MATLAB (https://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home). Statistical data analysis was performed using R programming language and RStudio software (Figure S1).

# 4 Results

In order to perform single cell expression analysis at different steps of neural differentiation, *E14*, *Nd* and *46C* mES cell lines were directed to neural differentiation in an adherent monolayer culture system, as described by Abranches *et al.*[52]. mES cells were harvested for different types of analysis at different time points, as schematized in Figure 4.




| FACS | All | All | All | All | All | All | All | All | All | All |
|---|---|---|---|---|---|---|---|---|---|---|
| Fix for smFISH | -- | -- | -- | *46C* | *46C* | *46C* | *46C* | *46C* | *46C* | -- |
| Immunofluorescence | -- | -- | -- | -- | -- | -- | -- | -- | All | -- |

**Figure 4-** Scheme of the monolayer neural differentiation protocol followed to direct neural differentiation of mES cells. The experimental procedures done at each time point are depicted (FACS, smFISH and immunostaining) as well as the cell line used for each procedure: All cell lines (*E14*, *Nd* and *46C*) or *46C* only.

## 4.1 Characterization of neural differentiation in mES cells

Characterization of mES cell neural differentiation in monolayer cultures was performed by analysis of both cell morphology (Figure S2) and expression of the Sox1-GFP reporter (Figure S3). Analysis of specific neural markers in rosette-like structures was also performed (Figure S4).

### *4.1.1 Cell morphology*

Cell morphology was monitored daily on an inverted bright field microscope. In pluripotency conditions, mES cells grew in characteristic mES organization with cells being arranged in clusters with few differentiated cells at the periphery (Figure S2-A; Day 0). Once cells are plated in the neural differentiation medium RHB-A, an increase in cell death (Figure S2-B) as well as a decrease in Fold Increase (FI) of the culture (Figure S2-C) was observed during the first days of neural differentiation, suggestive of a high medium selectivity. Also, the number of undifferentiated cell clusters decrease. When cells are replated at day 4 in RHB-A supplemented with basic FGF (bFGF), which is important for maintenance of neural progenitors, no undifferentiated cell clusters are formed and the differentiated cells are reorganized in rosettes-like structures. These display morphological and functional characteristics of the neural tube *in vivo* [52] (Figure S2-A; Day 6).

### *4.1.2 Dynamics of Nanog:VNP and Sox1-GFP reporters during monolayer differentiation*

In order to monitor efficiency of neural differentiation and exit from pluripotency, expression of Sox1-GFP and Nanog:VNP reporters was evaluated by FACS analysis, using *46C* and *Nd* cells, respectively. *E14* cells were used as a negative control for fluorescence (Figure S3). It was previously reported that *Sox1* expression is activated in proliferating neural progenitor cells, making *Sox1* an adequate gene to evaluate the neural differentiation effectiveness [58]. When measuring the levels of Sox1-GFP we observe an increase along the protocol reaching 80% GFP positive cells by day 4 (Figure S3), which is consistent with previous data [51,52] and confirms the efficiency of the protocol. Concerning the expression of Nanog:VNP, in pluripotency conditions (BMP4 LIF), 60% of the cells are Nanog:VNP positive, which is consistent with previous reports [31,43]. Once cells advance into neural fate, the percentage of Nanog:VNP decreases, as expected [24], reaching null levels at day 6 (Figure S3). It is also noteworthy that there is a small increase in Nanog:VNP levels at day 3 of neural differentiation, consistent with *in vivo* data [24]. Furthermore, this transient increase in NANOG levels was also confirmed at the mRNA level (section 4.2.1) corroborating previous studies [54,57,58].

### *4.1.3 Formation of rosette-like structures*

Efficiency of neural differentiation was further confirmed by assessing the expression of specific markers in rosette-like structures. Immunofluorescence was performed in cells at day 6 of neural differentiation to monitor expression of Sox1-GFP and N-Cadherin, which is present at apical adherent junctions, in order to confirm the organization of the neuroepithelial progenitors in rosette-like structures (Figure S4). These structures were proposed to mimic neural development *in* vivo, in which markers like N-Cadherin and ZO-1 are expressed in the apical domain of the embryonic neural tube[53]. Sox1-GFP expression was detected and proper cell organization in rosettes was identified, confirming efficient neural differentiation.

## 4.2 Single cell analysis of gene expression during early steps of neural differentiation

Population analyses provide insights on the mean behavior of a group of cells, however are less informative when trying to determine how individual cells coordinate specific gene expression when exiting pluripotency and entering neural commitment. Therefore, after confirming the efficiency of the neural monolayer differentiation protocol, *46C* cells were fixed to perform smFISH, a modified mRNA FISH method that allows a quantitative measurement of the number of mRNA molecules in single cells [59]. This method was used to quantify several mRNAs in *46C* cells, collected from different time points: Day 0, Day 1, Day 2, Day 3, Day 4 and Day 6 (Figure 4). In this single cell approach, it is possible to probe transcripts of three genes in each experiment. For all smFISH experiments performed, *Nanog* mRNA levels were

evaluated in order to study its putative transcriptional correlation with other genes, and also to prove the reproducibility of this approach.

For all analysed gene combinations, the distribution of the frequency of cells expressing different transcript levels, i.e. number of mRNA molecules, for each gene and experimental time point, was considered. The distribution shape could be useful to infer some relevant information about transcription state: a unimodal/Gaussian-like distribution, with the majority of cells expressing an average level of mRNA molecules, and few cells expressing less or more than the average, is suggestive of continuously active transcription with cells showing homogenous transcripts levels and therefore low cell-to-cell variability concerning the expression of that gene; long-tailed distributions, in which most cells express few transcripts, while a small number of cells displayed high number of transcripts, leading to a wide variety in the number of mRNA molecules present in the cells expressing that gene, and therefore to cell-to-cell heterogeneity. According to Raj and van Oudenaarden [59], this distribution is suggestive of a bursty-like transcription that occurs in infrequent but potent bursts that correlate with active and inactive states of transcription.

Furthermore, for each gene combination studied in this work, it was possible to establish correlations between the mRNA levels of each analysed gene in single cells. This enables us to group cells expressing similar transcript levels in subpopulations and infer their biological identity. In order to measure the statistical dependence of the two gene's transcription levels, Spearman correlation coefficient, r, was calculated. This value varies between -1 (negative correlation) and 1 (positive correlation). When r is high, cells with high expression of gene X are more likely to present high transcript levels of gene Y. The lowest the r value, the more likely is that cells with high expression of gene X have low transcript levels of gene Y.

For each analysed gene, a specific value of mRNA molecules per cell was defined as a threshold to distinguish cells expressing high transcript levels (number of mRNA molecules above threshold value) and low transcript levels (number of mRNA molecules below threshold value) (Table S6). This is important to distinguish cell populations in an active and inactive state of transcription, respectively. These values were assigned by observation of transcript number distributions in the experimental day of maximum expression for each genes, as described by Nair et al. [61].

### *4.2.1 Dismantling of the Pluripotency Gene Regulatory Network*

The first question I addressed concerned how the pluripotent gene regulatory network (measured by the expression of *Nanog*, *Oct4* and *Sox2*) is dismantled once cells are directed to neural lineage commitment. Cells in the pluripotent state (Day 0) and in the early steps of neural commitment (Day 1, Day 2 and Day 3) were collected, and smFISH was performed to detect *Oct4*, *Sox2* and *Nanog* transcripts. The distributions depicting the frequencies of cells

with a given expression value are shown in Figure 5. The correlations between the expression levels for each gene in single cells is analysed in Figure 6.

Concerning the distributions of mRNA molecules per cell for *Oct4*, *Sox2* and *Nanog*, at day 0, they are similar to published data[31,62], with *Oct4* presenting a unimodal, *Sox2* a bimodal and *Nanog* a long-tailed distribution. These distribution patterns suggest that, in pluripotency conditions, both *Nanog* and *Sox2* present a bursty-like transcription, as discussed in section 1.3. In addition, *Sox2* distribution reveals the existence of two cellular states within the analysed mES cell population (Figure 5): one expressing low numbers of *Sox2* mRNA molecules per cell, and another expressing higher levels of *Sox2* transcripts. The distribution of *Oct4* transcripts per cell, at day 0, suggests that its transcription occurs without bursts and with relatively small amount of noise, as indicated by its unimodal/Gaussian-like distribution[61]; however, this distribution could also be influenced by the relatively long half-live of *Oct4* mRNAs (Table S5) leading to the accumulation of mRNA molecules even when transcription had ceased, thereby masking an undergoing burst-like transcription [46].

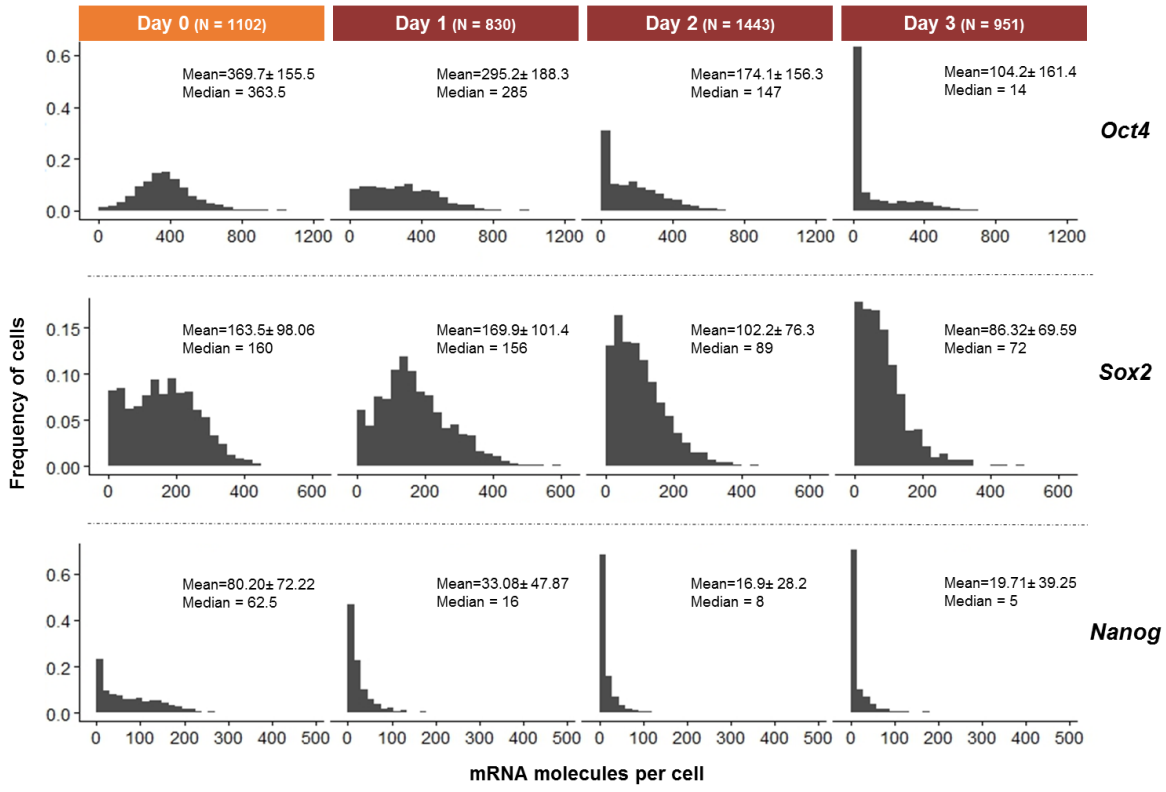



**Figure 5**- Histograms of the distribution of *Oct4*, *Sox2* and *Nanog* mRNA molecules per cell, for *46C* cells fixed at different time points. The number of analysed cells in each condition is shown in brackets. Mean±standard deviation and median values are shown to each cell population.

Upon neural commitment (Day 1 onwards), a decrease in the number of *Oct4* and *Nanog* transcripts per cell is observed, suggesting deactivation of the pluripotent gene regulatory network and gradual exiting from pluripotency. As for *Sox2*, it displays a small decrease in the number of transcripts per cell during neural commitment, in agreement with the

evidence that Sox2 is not only involved in the maintenance of the pluripotent state but also in neural commitment [22,24].

To analyse the co-expression of these genes in single mES cells, two-colour correlation dotplots were obtained (Figure 6). Concerning *Oct4-Sox2* correlation (Figure 6; top), it is possible to observe, at day 0, two cellular states: one in which cells with High-*Sox2* transcript levels express intermediate levels of *Oct4* transcripts (Figure 6; *2), and another in which low expression levels of *Sox2* are accompanied also by intermediate levels of *Oct4* expression (Figure 6; *1) . The first state is also characterized by the high levels of *Nanog* expression and represents the pluripotent state. The other state shows low levels of *Nanog* expression highlighting a population of pluripotent cells that is undergoing a negative phase of the fluctuations in *Nanog* expression [31].

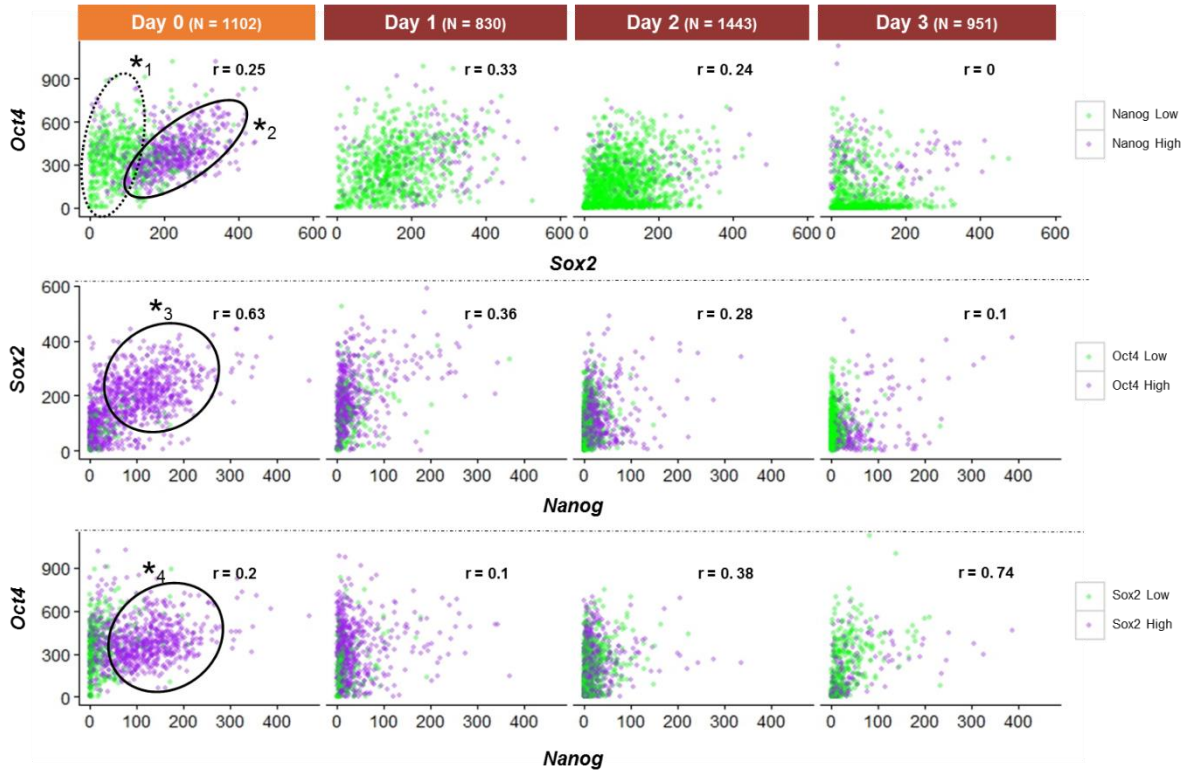



**Figure 6**- Two-colour dot plot graphs showing the correlations between the number of mRNA molecules of *Sox2*, *Oct4* and *Nanog* at different experimental days. r refers to the spearman correlation coefficient between the genes described in x and y axis.

Upon neural commitment, the number of mES cells at the pluripotent state decreases and, by day 3, only few cells remain pluripotent (high levels of *Oct4*, *Sox2* and *Nanog* mRNA molecules per cell). Along neural commitment, the correlation between the transcripts of *Oct4* and *Sox2* changes and, at day 3, the majority of cells expressing high levels of *Sox2* transcripts show low numbers of *Oct4* transcripts, meaning that when cells are exiting pluripotency (decrease of *Oct4* expression) into neural commitment, *Sox2* expression is maintained, highlighting the importance of its expression during commitment to neural fate. Concerning the correlation between *Nanog* and *Sox2* mRNA molecules per cell (Figure 6; middle), there is a

positive correlation at day 0 (r = 0.63), and cells with high levels of *Nanog* and *Sox2* expression have also high levels of *Oct4* transcripts (Figure 6; *3). This population disappears as cells commit to neural lineage, and this is more evident by the decrease in the number of *Nanog* mRNA molecules per cell. As for *Oct4* and *Nanog* correlation (Figure 6; bottom) the pluripotent population of mES cells is evident at day 0, expressing high levels of *Oct4*, *Nanog* and *Sox2* mRNA molecules per cell (Figure 6; *4). As expected, the number of mES cells expressing high levels of mRNAs from pluripotent genes decreases along neural commitment (Day 1 onwards). Interestingly, at day 3, a positive correlation between *Oct4* and *Nanog* mRNA levels(r= 0.74) is observed, although this value is mostly due to the high prevalence of cells with low expression levels of both genes that correspond to 73% of the analysed cells. Still, there is approximately 10% of cells expressing high levels of the two genes, and this might reflect a population of pluripotent epiblast-like cells that is known to be maintained in the embryo up to E8 [24].

These results confirm the higher expression of pluripotency genes (*Nanog*, *Oct4* and *Sox2*) at day 0, when mES cells are cultured in pluripotent conditions (BMP4 LIF). During neural lineage commitment, a decrease in the number of pluripotent cells is observed. This is more evident by following the decrease in *Oct4* and *Nanog* expression, the best correlate for the dismantling of the pluripotency network when mES cells commit to neural lineage.

For further analysis, a set of genes was selected: *Car2* and *Cldn6*, as stochastic genes, differentially expressed in pluripotent Low vs High NANOG cells; *Cdh2*, *Sox3* and *Crabp2* as genes known to be involved in neural commitment, *Fgf5* and *Dnmt3b* as genes upregulated in the postimplantation epiblast, and *T* as being an early marker of mesodermal commitment and when co-expressed with *Sox2* marking cells with NMp identity. Furthermore, transcripts from these genes have relatively short half-lifes[60] (Table S5), which facilitates the analysis of transcriptional dynamics along neural differentiation.

### *4.2.2 Expression of stochastic genes along neural differentiation*

*Car2* and *Cldn6* were selected based on their differential expression between Low-Nanog and High-Nanog cells (RNA-seq data; E. Abranches, unpublished data), with Low-Nanog cells displaying significantly higher expression levels of these genes compared to High-Nanog cells, suggesting that transcription of these genes occur when mES cells undergo “lineage priming”, becoming more susceptible to enter differentiation. *Car2* encodes for cytosolic carbonic anhydrase 2, and *Cldn6* encodes for the tight junction protein CLAUDIN6, involved in cell-cell adhesion. No known functions have been yet ascribed to these genes during pluripotency or entry into differentiation. In order to understand the transcriptional dynamics of these genes along neural commitment, *46C* cells fixed for smFISH at days 0, 1, 3 and 6 of the monolayer protocol were examined for *Car2*, *Cldn6* and *Nanog* transcripts. The distributions

depicting the frequencies of cells with a given expression value are shown in Figure 7. The correlations between the expression levels for each gene in single cells is analysed in Figure 8.

In the pluripotent state (day 0), transcription of all three genes is bursty (*Nanog* distribution not shown), as evidenced by the long-tailed shaped distributions (Figure 7). Upon neural commitment, *Car2* and *Cldn6* show an expression peak at day 1, followed by a gradual decrease in their transcription until day 6, when only few cells are expressing these genes. This is suggestive of a time window during neural commitment when cells are more permissive to explore the genome, probably as a result of a combination between the still "open chromatin" state from pluripotency with the downregulation of core pluripotency transcription factors that repress the expression of lineage-affiliated and other genes.

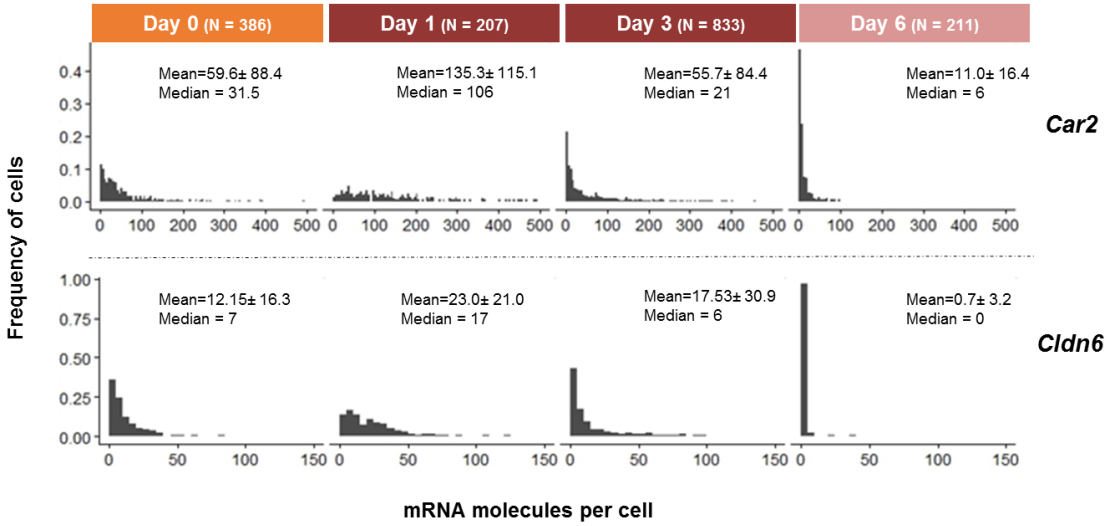



**Figure 7**- Histograms of the distribution of *Car2* and *Cldn6* mRNA molecules per cell, for *46C* cells fixed at different time points. The number of analysed cells in each condition is shown in brackets. Mean±standard deviation and median values are shown to each cell population.

When analysing gene expression correlations in single cells (Figure 8), it is possible to detect a state at day 0, in which the high *Car2*/*Cldn6*-expressing cells have a tendency to express low levels of *Nanog* transcripts (Figure 8; *), confirming its higher expression in pluripotent cells expressing low *Nanog* transcripts levels and suggesting a "priming" state in these cells (Figure 8; *). At day 1, when *Car2* and *Cldn6* show their highest expression levels (Figure 8), it is not possible to extract any relevant information from the correlation graphs. At day 3, the single cell correlations of *Car2*, *Cldn6* and *Nanog* expression differs from the previous days with approximately 70% of analysed cells expressing low transcript levels of all three genes, suggesting that the transcription of these gene is switched-off when cells enter differentiation. Concerning day 6, due to the low expression of these genes, expectedly, no correlation was found between their transcription.

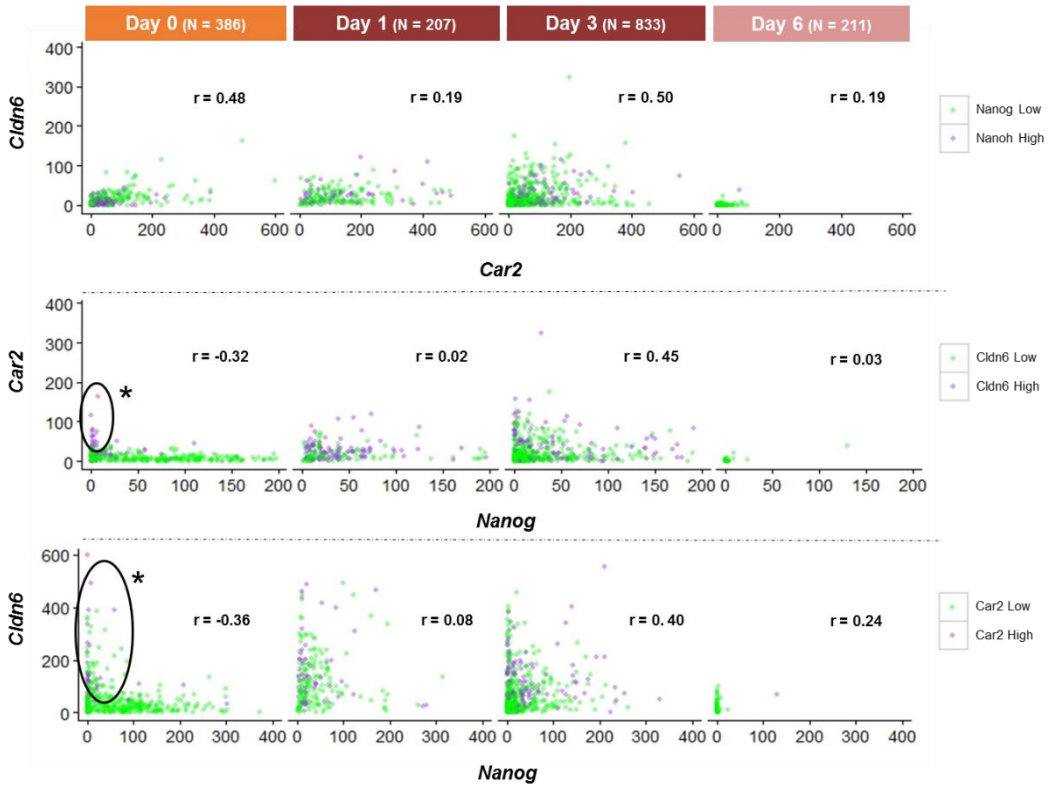



**Figure 8-** Two-colour dot plot graphs showing the correlations between the number of mRNA molecules of *Car2*, *Cldn6* and *Nanog* at different experimental days. r refers to the spearman correlation coefficient between the genes described in x and y axis.

In order to investigate whether there is a coordinated regulation of *Car2* and *Cldn6* in Low-*Nanog* cells, suggestive of a common upstream regulatory mechanism, the independence of these events (high expression of *Car2* and *Cldn6* transcripts) in cells with low levels of *Nanog* mRNA was evaluated (Table 2).

**Table 2**-Calculations of the percentage of Low *Nanog*-expressing cells that express high levels of *Car2* transcripts (A), high levels of *Cldn6* transcripts (B) or both (C). (D) represents the probability of Low *Nanog* cells express high transcript levels of both stochastic genes, if (A) and (B) are considered independent events. These calculations were performed for each experimental day.

| | In Nanog Low cells | | | |
|---|---|---|---|---|
| | Car2 High (%) (A) | Cldn6 High (%) (B) | Car2 High and Cldn6 High (%) (C) | P (Car2 High and Cldn6 High) (%) if (A) and (B) are independet events (D) (D) = (A)x(B) |
| Day 0 | 29 | 15 | 10 | 4 |
| **Day 1** | 52 | 28 | **17** | **15** |
| **Day 3** | 15 | 16 | 7 | 2 |
| **Day 6** | 0 | 0 | 0 | 0 |

For that, I calculated for each experimental day the frequency of cells expressing only high levels of *Car2* transcripts (value A, Table 2), the frequency of cells expressing only high levels of *Cldn6* transcripts (value B, Table 2) and the frequency of cells expressing high levels of both *Car2* and *Cldn6* (value C, Table 2) considering only the sub-population of Low *Nanog*-expressing cells (Table 2). When two events are independent, the probability of them occurring together is equal to the multiplication of the probability of each separated occurrence, and this was also calculated (value D, Table 2). The results indicate that the High-*Car2* and High-*Cldn6*

expression states, at day1, could be considered independent events, because the (D) value is very similar to (C) value. For day 6, it is not possible to evaluate the events' independence because no High *Car2* or *Cldn6* expression is observed in Low *Nanog*-expressing cells. Concerning days 0 and 3, (D) and (C) values are different, suggesting that expression of high *Car2* and *Cldn6* in Low-*Nanog* cells are dependent events, proposing the existence of a common molecular mechanism promoting high expression of both genes.

Overall, these results confirm the differential expression of *Car2* and *Cldn6* in *Nanog*-expressing cells at day 0, with cells that have high expression levels of *Car2* and *Cldn6* presenting simultaneously low levels of *Nanog* transcripts. Conversely, High *Nanog*-expressing cells present low numbers of *Car2* and *Cldn6* transcripts. Expression of these stochastic genes peaks at Day 1, decreasing in the following days of neural commitment. Additionally, this high expression of *Car2* and *Cldn6* is independent, suggesting that it is not driven by a common pathway. This supports the previous hypothesis that mES cells "explore" the whole-transcriptome in the early steps of differentiation, especially at Day 1, when still exiting from pluripotency.

### *4.2.3 Expression of lineage-affiliated genes along neural differentiation*

My next aim was to analyse the expression of lineage-affiliated genes during different time points of neural commitment. The selected genes have different developmental roles: *Cdh2*, *Sox3* and *Crapb2* are genes involved in neural commitment; *Fgf5* and *Dnmt3b* are known genes to be upregulated during early differentiation *in vitro* mimicking their *in vivo* expression in the mouse postimplantation epiblast; and joint expression of *T* and *Sox2* mark bipotent NMp's.

#### *4.3.3.1 Neural genes: Cdh2, Sox3 and Crabp2*

*Cdh2* encodes for N-Cadherin, which is involved in neuroectoderm formation and is important for nervous system development[63], as well as for formation of rosette-like structures during *in vitro* neural differentiation (Figure S4). *Sox3* encodes for a transcription factor of the Sox (SRY-related HMG box) family, and belongs to the SOX B1 subgroup, which also includes SOX1 and SOX2. SOX3 expression in the mouse has been observed in the epiblast and extraembryonic ectoderm of the egg cylinder, becoming restricted to the prospective neural plate at the onset of gastrulation[64] and it has been shown to be a central player in the maintenance of neural stem cells[65]. *Crabp2* encodes for a cellular retinoic binding protein, involved in the transport of retinoic acid to the nuclear retinoic acid receptors, and its expression was found increased in the transition to a proliferative neurogenic population state during monolayer neural differentiation[52].

In order to understand the transcriptional dynamics of these genes along neural differentiation, *46C* cells fixed for smFISH at days 0, 1, 3, 4 and 6 of the monolayer protocol

were examined for their levels of *Cdh2*, *Sox3*, *Crabp2* and *Nanog* transcripts. The distributions depicting the frequencies of cells with a given expression value are shown in Figure 9. The correlations between the expression levels for each gene in single cells is analysed in Figure10 and 11.

For *Cdh2* (Figure 9; up), it is possible to observe a continuous increase in the frequency of cells expressing high *Cdh2* mRNA levels, whereas no expression could be detected in pluripotent cells (day 0), confirming that *Cdh2* expression is a good marker for neural differentiation.

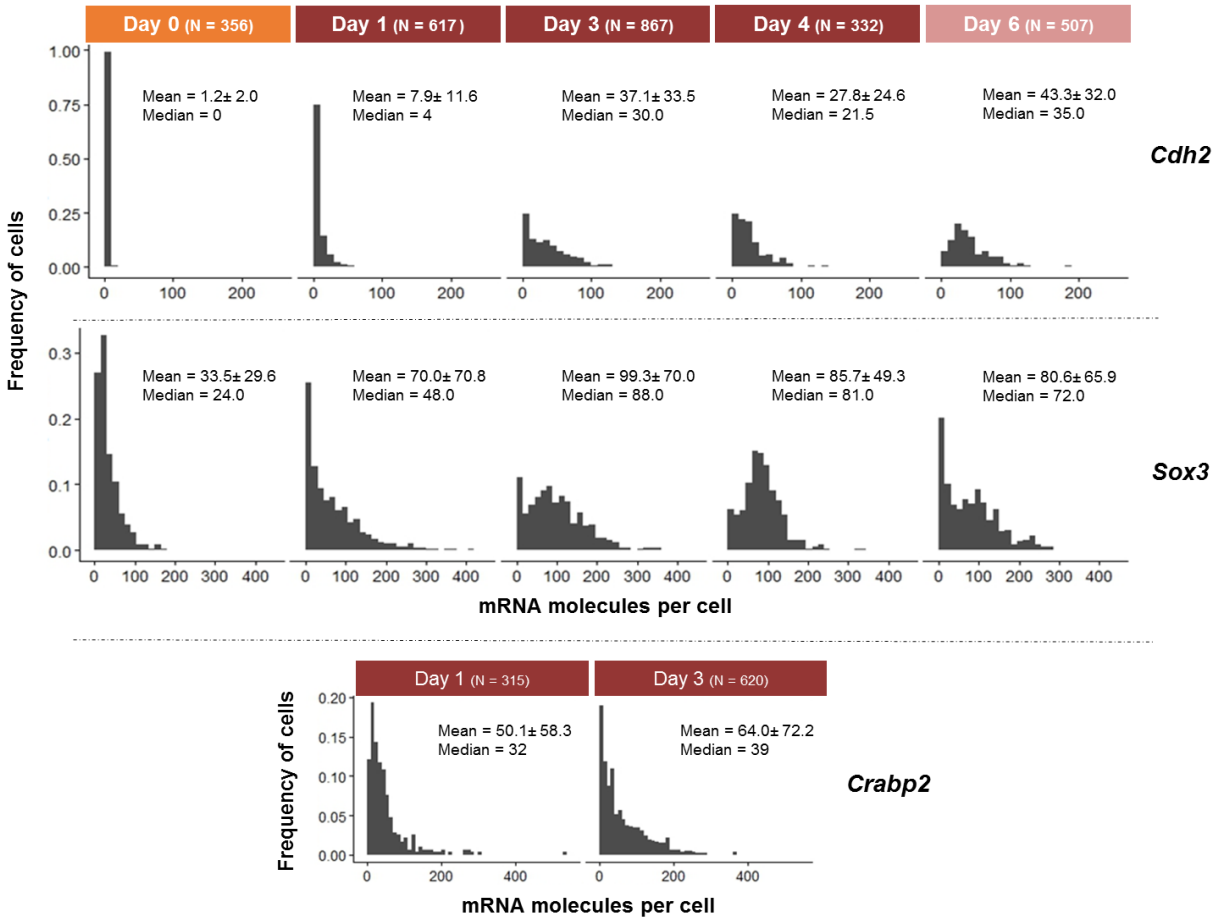



**Figure 9**- Histograms of the distribution of *Cdh2*, *Sox3* and *Crabp2* mRNA molecules per cell, for *46C* cells fixed at different time points. The number of analysed cells in each condition is shown in brackets. Mean±standard deviation and median values are shown to each cell population.

For *Sox3* (Figure 9; middle), some mES cells at day 0 already express significant number of *Sox3* mRNA molecules, which is in agreement with other studies that found *Sox3* to be expressed in the epiblast, before commitment to any germ layer[64] . Concerning *Sox3* mRNA levels per cell during neural commitment (Day 1 onwards), a peak of *Sox3* expression occurs at days 3 and 4, in agreement with its known role in the early steps of neural commitment.

Regarding *Crabp2* (Figure 9; bottom), it was only possible to study its expression at days 1 and 3 of the monolayer neural differentiation. At day 1, some cells already express

considerable number of *Crabp2* mRNA molecules, suggesting that, *Crabp2* is also important, during neural commitment, as confirmed by the increase in its expression at day 3.

Concerning the single cell correlations observed for the expression of these genes (Figure 10 and 11), there is a noticeable correlation between *Sox3 and Cdh2* expression (Figure 10; up) in the first days of neural differentiation (Days 1, 3 and 4) with the majority of the cells that co-express these genes showing low numbers of *Nanog* mRNA molecules per cell representing a committed population to neural fate.

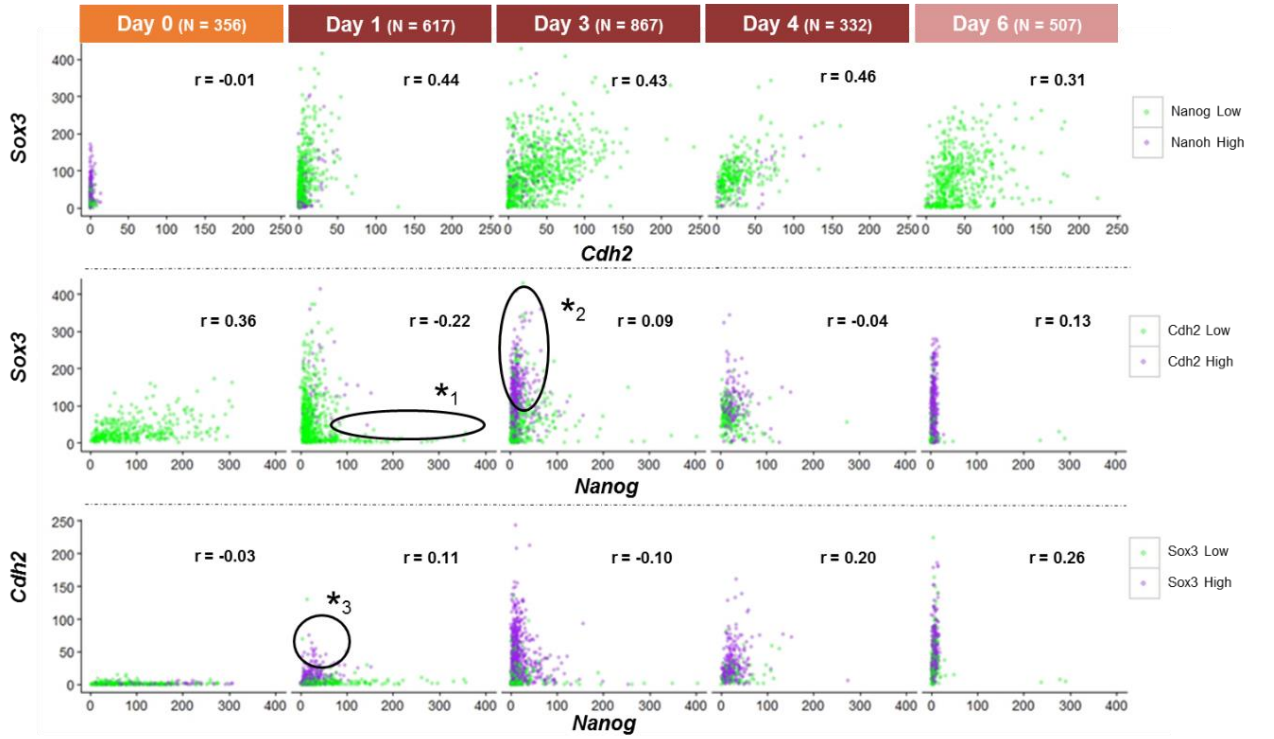



**Figure 10**- Two-colour dot plot graphs showing the correlations between the number of mRNA molecules of *Cdh2*, *Sox3* and *Nanog* at different experimental days. r refers to the spearman correlation coefficient between the genes described in x and y axis.

For *Sox3* and *Nanog* (Figure 10; middle), an anti-correlation (r= -0.22) is observed at day 1, with High-*Nanog* cells containing low numbers of *Sox3* mRNA molecules (Figure 10; *1). From day 3 onwards, cells with high *Cdh2* expression (Figure 10; *2) also express high levels of *Sox3* mRNA and simultaneously low levels of *Nanog* transcripts, highlighting the cell population committed to neural fate.

For *Cdh2* and *Nanog* correlations (Figure 10; down), high *Cdh2* expressing cells, not present at day 0, simultaneously express high transcript levels of *Sox3* but low transcript levels of *Nanog* (Figure 10; *3) at day 1. Following neural differentiation timeline, a decrease in the frequency of cells expressing high transcript levels of *Nanog* is evident, as well as an increase in the frequency of cells co-expressing high levels of *Cdh2* and *Sox3*, supporting once more the exit from pluripotency (decrease in *Nanog* expression) following neural commitment (increase in *Cdh2* and *Sox3* expression).

Concerning *Crabp2*-*Cdh2* correlation (Figure 11; up), an increase is observed from day 1 to day 3, similar to that observed for the other neural genes *Cdh2* and *Sox3* (Figure 10; up). For *Crabp2* and *Nanog* (Figure 11; down), an anti-correlation in both days is observed, being more evident at day 3, with high *Crabp2* expressing cells showing low levels of *Nanog* expression. In this state, high *Cdh2* expressing cells show high transcription levels of *Crabp2* and low number of *Nanog* mRNA molecules (Figure 11; *), supporting the increased frequency of cells entering neural commitment.

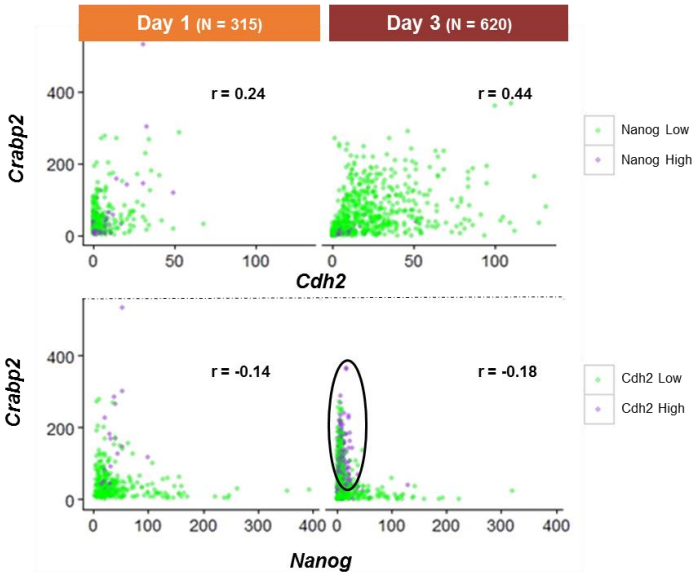



**Figure 11**- Two-colour dot plot graphs showing the correlations between the expression of *Cdh2*, *Crabp2* and *Nanog* at different days. r refers to the spearman correlation coefficient for each correlation.

These results indicate that *Crabp2* and *Sox3* transcription, unlike *Cdh2*, is not exclusive of cells that enter neural commitment, but are also expressed in pluripotent cells. In addition, expression of high mRNA levels of *Cdh2*, *Crabp2* and *Sox3* tends to occur in cells with low *Nanog* expression (from day 1 onwards, 80 to 100% of high *Cdh2*/*Sox3*-expressing cells and 70 to 100% of high Cdh2/Crabp2-expressing cells present low levels of *Nanog* transcripts) revealing the existence of two simultaneous processes when cells transit from pluripotency to lineage commitment: decrease of *Nanog* expression and upregulation of lineage-affiliated genes.

#### *4.3.3.2 Postimplantation Epiblast genes: Fgf5 and Dnmt3b*

When mouse embryos undergo implantation, several morphologic and transcriptomic alterations occur. Molecular landmarks of this transition are: dismantling of the pluripotency network and upregulation of postimplantation epiblast markers like *Fgf5*, *Oct6* and *Otx2*, and of the *de novo* methyltransferases *Dnmt3a1* and *Dnmt3b* [16]. This increased expression of postimplantation epiblast markers was also reported in the first days of different *in vitro* mES differentiation systems [52,54,57,58]. In order to monitor the expression of these postimplantation epiblast markers during monolayer neural differentiation, cells fixed for smFISH at days 1 and

3 of the monolayer protocol were analysed for *Fgf5*, *Dnmt3b* and *Nanog* transcripts. The distributions depicting the frequencies of cells with a given expression value are shown in Figure 12A. The correlations between the expression levels for each gene in single cells is analysed in Figure 12B.

For *Fgf5* and *Dnmt3b*, a noticeable increase in the frequency of cells expressing high levels of both mRNAs can be observed from day 1 to day 3 of the monolayer neural differentiation protocol (Figure 12A).Concerning the single cell correlations observed for the expression of *Fgf5*, *Dnmt3b* and *Nanog* expression (Figure 12B), an increase in the correlation between all gene combinations (*Fgf5-Dnmt3b*, *Nanog-Fgf5* and *Nanog-Dnmt3b*) is observed. This is supported by a significant increase in the Spearman correlation coefficient, r, from day 1 to day 3, for all comparisons. This increased correlation between the expression of *Nanog* and *Fgf5* and *Dnmt3b* most likely reflects the transient increase of *Nanog* expression reported at day 3 of monolayer neural differentiation (Figure 5), coinciding with the upregulation in the expression of the epiblast markers.

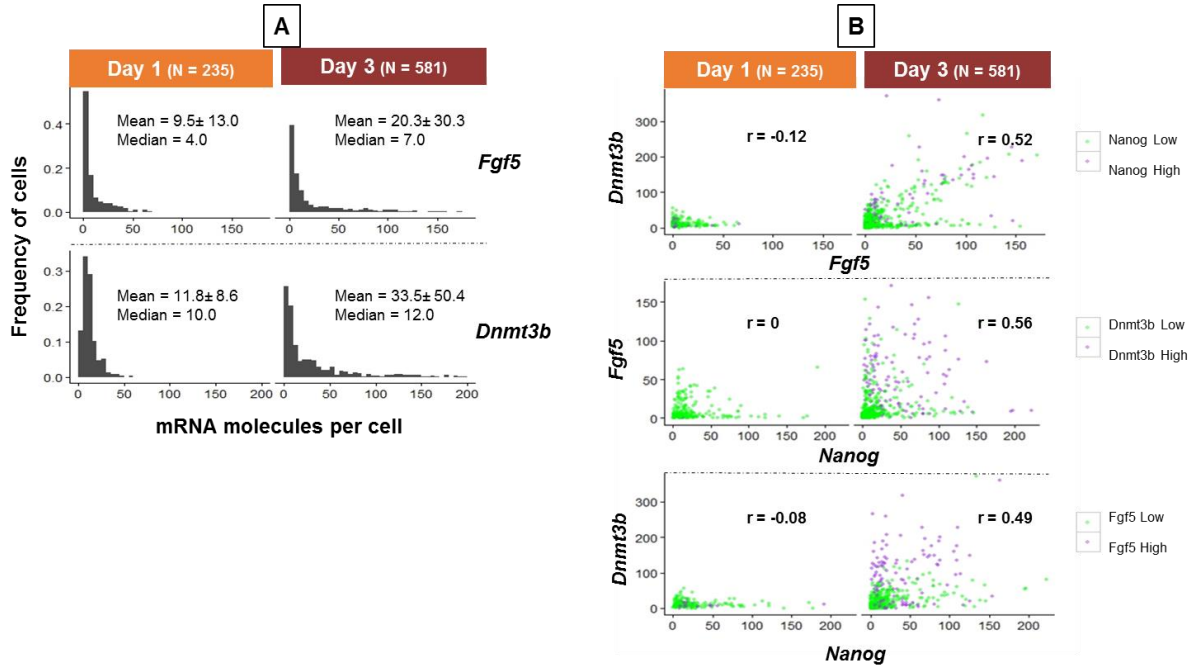



**Figure 12**- (**A**) Histograms of the distribution of *Fgf5* and *Dnmt3b* mRNA molecules per cell, for *46C* cells fixed at different time points. The number of analysed cells in each condition is shown in brackets. Mean±standard deviation and median values are shown to each cell population; (**B**) Two-colour dot plot graphs showing the correlations between the number of mRNA molecules of *Fgf5*, *Dnmt3b* and *Nanog* at different experimental days. r refers to the spearman correlation coefficient for each correlation.

These results indicate that at day 3 of this *in* vitro differentiation system is possible to capture cells in an epiblast-like stage, marked by transient reactivation of *Nanog* and upregulation of postimplantation epiblast markers, like *Fgf5* and *Dnmt3b*. This observation is concordant with other reports [54,58].

#### *4.3.3.3 NMp gene's: T and Sox2*

In order to evaluate the generation of NMp-like cells during this *in vitro* neural differentiation system, the expression of *T* and *Sox2* was assessed at days 1 and 3 of the differentiation protocol. The distributions depicting the frequencies of cells with a given expression value are shown in Figure 13A. The correlations between the expression levels for each gene in single cells is analysed in Figure 13B.

As shown in Figure 13A, the number of *T* mRNA molecules per cell is low at all analysed time points. However, there are few cells expressing high levels of *T* mRNA, at day 3. No significant correlation is found between *T* and *Sox2* and *T* and *Nanog* expression levels (Figure 13B). Thus, cells coexpressing high levels of T and Sox2 mRNA molecules are rare at both time points, mainly due the residual transcription of *T*.

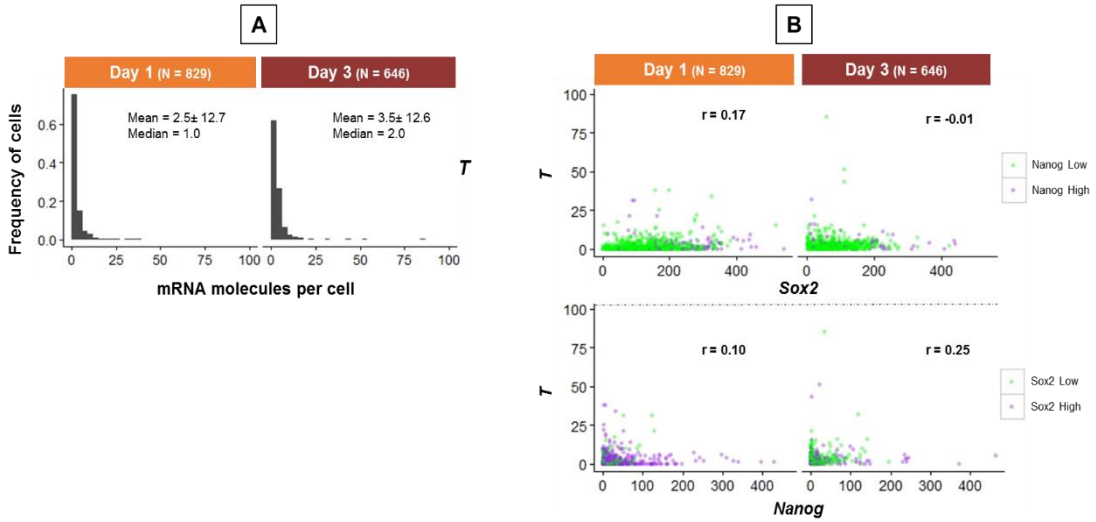



**Figure 13**- (**A**) Histogram of the distributions of *T* mRNA molecules per cell, for *46C* cells fixed at different time points. The number of analysed cells in each condition is shown in brackets. Mean±standard deviation and median values are shown for each cell population; (**B**) Two-colour dot plot graphs showing the correlations between the number of mRNA molecules of *T*, *Sox2* and *Nanog* at different experimental days. r refers to the spearman correlation coefficient for each correlation.

These results confirm the assumption that neural differentiation in RHB-A media, like in N2B27, favours anterior neural development, and, in the absence of "posteriorizing" signals, no posterior neural induction occurs and no NMp-like cells are observed.

# 5 Discussion and Conclusion

In this study, my aim was to i) understand how is the core pluripotency network (*Nanog*, *Oct4* and *Sox2*) dismantled along neural differentiation, ii) understand the transcription dynamics of stochastic and lineage-affiliated genes along neural differentiation and iii) assess the presence of cells co-expressing *T* and *Sox2* (NMp's markers) along neural differentiation.

Concerning the dismantling of pluripotent gene regulatory network, my results confirm the high and heterogeneous single cell expression of *Nanog*, *Sox2* and *Oct4* in pluripotency conditions (BMP4 LIF), also shown in population analyses studies. My results show that two cellular subpopulations exist in pluripotency conditions: one undifferentiated subpopulation with high expression of pluripotency genes, and another with intermediate levels of *Oct4* and low levels of *Sox2* and *Nanog* mRNAs. The variable expression of *Nanog* mRNA levels in pluripotency conditions was expected and previously described [31]. However, heterogeneous expression of *Sox2* in the pluripotent state is intriguing. One possible explanation, is that BMP4, used in combination with LIF to maintain cells in a pluripotent state, might inhibit *Sox2* expression. This inhibitory effect was reported in chick embryos, in which misexpression of BMP4 in the prospective neural plate inhibits *Sox2* and *Sox3* expression[66] and by Wakamatsu et al. [67] that showed an inhibitory effect of BMP4 on *Sox2* expression in avian neural crest cells. However, this hypothesis does not explain the cell-to-cell variability of *Sox2* expression, and further studies are needed to confirm this putative effect in *Sox2* expression in mES cells. My results indicate that, upon neural commitment, a decrease in the number of *Oct4* and *Nanog* transcripts per cell is observed, suggesting deactivation of the pluripotent gene network and gradual exiting from pluripotency, as previously described during *in vitro* neural differentiation [54], and during mouse embryogenesis [24]. However, at day 3 of the neural differentiation protocol, approximately 10% of cells express high levels of the two genes, which might reflect the existence of a population of pluripotent epiblast-like cells, which is known to be maintained in the embryo up to E8 [24]. *Sox2* expression pattern is different and, unlike *Nanog* and *Oct4,* displays a small decrease in the number of transcripts per cell during neural commitment, being therefore expressed during neural lineage commitment. This confirms the importance of *Sox2* expression not only in the maintenance of the pluripotent state but also in neural development.

Concerning the expression of stochastic genes, *Car2* and *Cldn6,* in the pluripotent state, transcription is bursty, as evidenced by the long-tailed shaped distributions. In this state, high *Car2*/*Cldn6*-expressing cells have a tendency to express low levels of *Nanog* transcripts, suggesting a "priming" state in these cells. Upon neural commitment, *Car2* and *Cldn6* show an expression peak at day 1, followed by a gradual decrease in their transcription until day 6. This is suggestive of a time window during neural commitment when cells are more permissive to explore the genome, probably as a result of a combination between the still "open chromatin" state from pluripotency with the downregulation of core pluripotency transcription factors that

repress the expression of lineage-affiliated and other genes. When testing the independence of High *Car2* and *Cldn6* expression in Low-Nanog cells during neural differentiation, to investigate whether there is a coordinated regulation of these genes, I was able to observe that the High-*Car2* and High-*Cldn6* expression, at day1, could be considered independent events, in opposition to days 0 and 3, suggesting that, at day 1, transcription of these stochastic genes is not driven by a common pathway, supporting the previous hypothesis that mES cells "explore" the whole-transcriptome in the early steps of differentiation, when exiting from pluripotency.

When I analysed the single cell expression of neural-affiliated genes during neural differentiation, *Cdh2*, *Sox3* and *Crabp2* presented a bursty-like transcription during the first days of differentiation, as shown by the long-tailed distribution of the frequency of cells expressing different mRNA levels, providing a potential mechanism for "priming" of cellular fate decisions. However, a less noisy transcription is present later in the differentiation system, with *Cdh2* and *Sox3* showing a unimodal-like distribution at days 6 and 4, respectively, occurring a switch from bursty-like to constitutive transcription of these neural genes. The same cannot be assumed for *Crabp2* because its single cell expression was only assessed at days 1 and 3 of the neural differentiation protocol. These observations fit a model [68] that states that intrinsic lineage choices depend on two sequential events: a) activation of lineage-affiliated transcription factors and b) their expression above a certain level. Therefore, during the first days of differentiation, neural genes *Cdh2* and *Sox3* are expressed in infrequent bursts and the frequency of cells expressing high transcript levels of these genes is consequently low. However, latter in neural differentiation, these genes are expressed in a constitutive manner, above a certain level, turning cells irreversibly into neural lineage fate.

Concerning the single cell expression of postimplantation epiblast markers, *Fgf5* and *Dnmt3b*, an increased expression was observed at day 3 of neural differentiation. This increase is correlated with an upregulation of *Nanog* expression. This defines a cell subpopulation co-expressing *Fgf5*. *Nanog* and *Dnmt3b*, most likely representing a time window when cells acquire an epiblast-like population, also reported in other *in vitro* differentiation studies[54,57,58].

Altogether, these results suggest that the initial days of differentiation are a "lineage priming" period characterized by the disassembling of core pluripotency network and increased expression of lineage-affiliated genes, accompanied by random expression of stochastic genes with no developmental role. This period precedes a definitive neural commitment, with the constitutive expression of genes involved in neuroectodermal development (*Cdh2* and *Sox3*). This point of view meets Zhang et al.[69] model arguing that neural differentiation of mES cells occurs in two stages: the first involves the transition from pluripotent mES cells to a "primed" state similar to mEpiS cells, and the second involves the transition to neural progenitor cells.

When searching for the existence of putative NMp-like cells, assessing the single cell coexpresion of *T* and *Sox2* transcripts, only rare cells were found to coexpress *T* and *Sox2*.

Unlike others reports [57], also studying gene expression during mES neural differentiation, the transcription of *T* was residual. One possible explanation is that RHB-A, the neural differentiation medium used in my studies, is highly selective when compared to N2B27, as showed by Abranches et al.[52], inhibiting the survival of cells primed for a non-neural lineage. Besides that, mES cells differentiating in RHB-A, like in N2B27, express high levels of anterior neural genes but lack expression of posterior neural genes [70,71], present in posterior tissues originated from NMp cells in the posterior epiblast in E8.5 mouse embryos [32]. Therefore, it would be interesting to investigate single cell expression of pluripotent, stochastic and lineage-affiliated genes when "posteriorizing" signals, agonists of Wnt and FGF signaling, are added to the culture media. In this differentiation protocol, both anterior and posterior neural cells (including NMp-like cells) are expected to arise and therefore this could provide a more comprehensive knowledge of neural differentiation.

In future work, it will be important to extend the single cell expression analysis to more experimental days of the monolayer neural differentiation protocol, to have a complete study of single cell transcriptional dynaamics for all analysed genes through the neural differentiation protocol, as well as include more mesoendodermal and neural genes (anterior and posterior) in the set of analysed genes to provide a broader knowledge of neural differentiation. Comparative study of single cells from mouse embryos will also be significant to understand to what extent our *in vitro* differentiation system correlates and mimics transcription in embryonic development.

Concerning the assignment of threshold values used in this study to distinguish low and high expressing values for each gene, and consequently distinguish inactive from active transcribing cells. For a clearer definition of these values, smFISH using intronic probes should be performed. In this approach, cells are incubated with probes targeting introns of the nascent mRNAs of the gene of interest. Here, only 4 different outcomes could be obtained: no nascent mRNA is detected and therefore the gene is not being transcribed; one nascent mRNA is detected, meaning that transcription is occurring only from one allele; two nascent mRNA alleles are detected suggesting that transcription is occurring from both alleles and four mRNA alleles are detected meaning that the cell is at G2/mitosis and transcription is activated in both alleles. From this experiment, cells with no detected nascent mRNAs are not transcribing the studied gene, in contrast with cells that present 1, 2 or 4 nascent mRNAs, due to active transcription. When correlating the number of nascent mRNAs (using intronic probes) and the number of mature mRNAs (using exonic probes) for the same gene, it should be possible to analyse the number of mature mRNAs in cells where transcription of the studied gene is inactive and the number of mature mRNAs in cells where the gene is actively transcribed. Comparing these values, it should be possible to extract an intermediate number of transcripts that could distinguish cells that are in an inactive or active transcription mode concerning the studied gene.

# 6 References


1. Bedzhov, I., Graham, S. J., Leung, C. Y. & Zernicka-Goetz, M. Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369,** (2014).
2. Ohnishi, Y. *et al.* Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* **16,** 27–37 (2014).
3. Morris, S. A. *et al.* Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 6364–9 (2010).
4. Sheng, G. Epiblast morphogenesis before gastrulation. *Dev. Biol.* **401,** 17–24 (2015).
5. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–6 (1981).
6. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **78,** 7634–8 (1981).
7. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16,** 516–28 (2014).
8. Loh, K. M., Lim, B. & Ang, L. T. Ex uno plures: molecular designs for embryonic pluripotency. *Physiol. Rev.* **95,** 245–95 (2015).
9. Kinoshita, M. How are pluripotent cells captured in culture? *Reprod. Med. Biol.* **14,** 85–98 (2015).
10. Rossant, J. Stem cells and early lineage development. *Cell* **132,** 527–31 (2008).
11. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4,** 487–92 (2009).
12. Hackett, J. A. & Surani, M. A. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell* **15,** 416–30 (2014).
13. Donaldson, DD, Wongf, GG, Moreau, J & Stahlt, M. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. … (1988). at <http://www.genomics.princeton.edu/tank/pdf-publications/Nature%20336,%20677-680%20(1988)..pdf>
14. Williams, RL, Hilton, DJ, Pease, S & Willson, TA. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* (1988). at <http://depts.washington.edu/iscrm/education/Conj530Reading/Ware10.pdf>
15. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12,** 2048–60 (1998).
16. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369,** (2014).
17. Ying, QL, Nichols, J, Chambers, I & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* (2003). doi:10.1016/S0092-8674(03)00847-X
18. Ying, Q.-L. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).
19. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132,** 532–6 (2008).
20. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379–91 (1998).
21. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–42 (2003).
22. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126–40 (2003).
23. Palmieri, SL, Peter, W, Hess, H & Schöler, HR. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Developmental biology* (1994). at <http://www.sciencedirect.com/science/article/pii/S0012160684713121>
24. Osorno, R. *et al.* The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* **139,** 2288–98 (2012).

25. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24,** 372–6 (2000).
26. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643–55 (2003).
27. Dietrich, J.-E. & Hiiragi. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134,** 4219–4231 (2007).
28. Boroviak, T. & Nichols, J. The birth of embryonic pluripotency. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369,** (2014).
29. Yamaguchi, S, Kimura, H, Tada, M & Nakatsuji, N. Nanog expression in mouse germ cell development. *Gene Expression …* (2005). at <http://www.sciencedirect.com/science/article/pii/S1567133X05000219>
30. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–4 (2007).
31. Abranches, E. *et al.* Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* **141,** 2770–9 (2014).
32. Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. Neuromesodermal progenitors and the making of the spinal cord. *Development* 1–12 (2015). doi:10.1242/dev.119768
33. Masui, S, Nakatake, Y, Toyooka, Y & Shimosato, D. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell …* (2007). doi:10.1038/ncb1589
34. Chew, J.-L. L. *et al.* Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell. Biol.* **25,** 6031–46 (2005).
35. Zhao, S, Nichols, J, Smith, AG & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Molecular and Cellular Neuroscience* (2004). at <http://www.sciencedirect.com/science/article/pii/S104474310400185X>
36. Shi, W *et al.* Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. *Journal of Biological …* (2006). doi:10.1074/jbc.M601811200
37. Gagliardi, A, Mullin, NP, Tan, ZY & Colby, D. A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *The EMBO …* (2013). at <http://emboj.embopress.org/content/32/16/2231.short>
38. Parfitt, D.-E. E. & Shen, M. M. From blastocyst to gastrula: gene regulatory networks of embryonic stem cells and early mouse embryogenesis. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369,** (2014).
39. Singh, A. M., Hamazaki, T., Hankowski, K. E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25,** 2534–42 (2007).
40. Torres-Padilla, M.-E. E. & Chambers, I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development* **141,** 2173–81 (2014).
41. Li, Y. *et al.* Gene expression profiling reveals the heterogeneous transcriptional activity of Oct3/4 and its possible interaction with Gli2 in mouse embryonic stem cells. *Genomics* **102,** 456–67 (2013).
42. Smith, A. Nanog heterogeneity: tilting at windmills? *Cell Stem Cell* (2013). doi:10.1016/j.stem.2013.06.016
43. Abranches, E., Bekman, E. & Henrique, D. Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of Nanog expression. *PLoS ONE* **8,** e59928 (2013).
44. Kalmar, T, Lim, C, Hayward, P & Muñoz-Descalzo, S. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS biology* (2009). doi:10.1371/journal.pbio.1000149
45. Arias, AM & Brickman, JM. Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Current opinion in cell biology* (2011). doi:10.1016/j.ceb.2011.09.007
46. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4,** e309 (2006).
47. Elowitz. Stochastic Gene Expression in a Single Cell. *Science* **297,** 1183–1186 (2002).
48. Chang, H., Hemberg, M., Barahona, M., Ingber, D. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* (2008). doi:10.1038/nature06965

49. Nieuwkoop, PD. Activation and organization of the central nervous system in amphibians. Part I. Induction and activation. *Journal of Experimental Zoology* (1952). doi:10.1002/jez.1401200102
50. Tzouanacou, E. *et al.* Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Developmental Cell* (2009). doi:10.1016/j.devcel.2009.08.002
51. Ying, Q.-L. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21,** 183–6 (2003).
52. Abranches, E. *et al.* Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo. *PLoS ONE* **4,** e6286 (2009).
53. Chenn, A., Zhang, Y. A., Chang, B. T. & McConnell, S. K. Intrinsic polarity of mammalian neuroepithelial cells. *Mol. Cell. Neurosci.* **11,** 183–93 (1998).
54. Trott, J. & Martinez Arias, A. Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. *Biol Open* **2,** 1049–56 (2013).
55. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5,** 877–9 (2008).
56. Pevny, LH, Sockanathan, S & Placzek, M. A role for SOX1 in neural determination. … (1998). at <http://dev.biologists.org/content/125/10/1967.short>
57. Nichols, J, Kranz, A, Stewart, AF & Smith, A. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* (2012). at <http://www.sciencedirect.com/science/article/pii/S0092867412004096>
58. Turner, DA, Trott, J, Hayward, P, Rué, P & Arias, AM. An interplay between extracellular signalling and the dynamics of the exit from pluripotency drives cell fate decisions in mouse ES cells. *Biology open* (2014). at <http://bio.biologists.org/content/early/2014/06/13/bio.20148409.short>
59. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135,** 216–26 (2008).
60. Sharova, L. V. *et al.* Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA research* **16,** 45–58 (2009).
61. Nair, G., Abranches, E., Guedes, A., Henrique, D. & Raj, A. Heterogeneous lineage marker expression in naive embryonic stem cells is mostly due to spontaneous differentiation. *Sci. Rep.* **5,** (2015).
62. Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55,** 319–31 (2014).
63. Redies, C. Cadherins in the central nervous system. *Progress in neurobiology* **61,** 611–648 (2000).
64. Wood, HB & Episkopou, V. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mechanisms of development* (1999). at <http://www.sciencedirect.com/science/article/pii/S0925477399001161>
65. Bylund, M, Andersson, E, Novitch, BG & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1–3 activity. *Nature neuroscience* (2003). doi:10.1038/nn1131
66. Linker, C. & Stern, C. D. Neural induction requires BMP inhibition only as a late step, and involves signals other than FGF and Wnt antagonists. *Development* **131,** 5671–5681 (2004).
67. Wakamatsu, Y, Endo, Y & Osumi, N. Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. *Developmental …* (2004). doi:10.1002/dvdy.10498
68. Nimmo, R., May, G. & Enver, T. Primed and ready: understanding lineage commitment through single cell analysis. *Trends in cell biology* (2015).
69. Zhang, K *et al.* Distinct functions of BMP4 during different stages of mouse ES cell neural commitment. … (2010). doi:10.1242/dev.049494
70. Gouti, M. *et al.* In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. (2014). doi:10.1371/journal.pbio.1001937
71. Turner, D. A. *et al.* Wnt/ -catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development* **141,** 4243–4253 (2014).

# 7 Supplementary Information

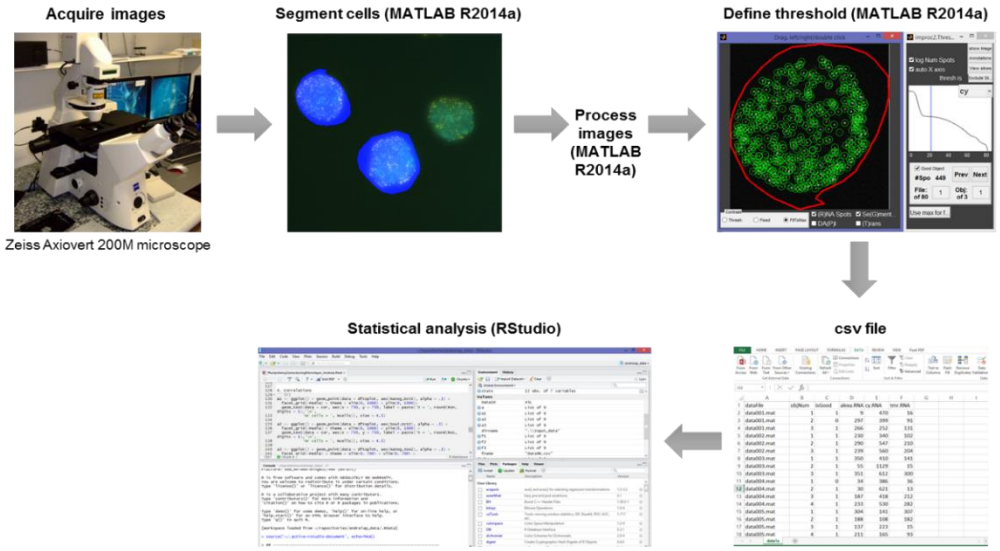



**Figure S1-** Diagram representative of the analysis flow of each smFISH experiment.

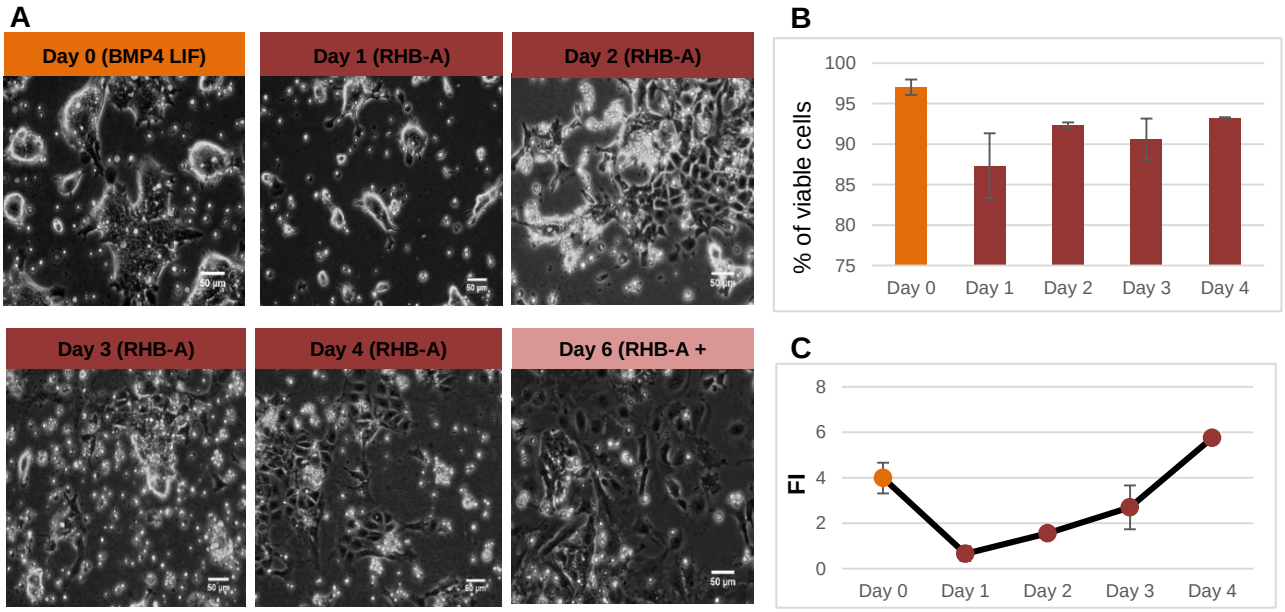



**Figure S2**- **A)** Bright field representative images of *46C* mES cells cultured along monolayer neural differentiation. All cell lines showed typical morphology at each time point and are able to reorganize in neuroepithelial rosettes at day 6. (Scale bar = 50µm); **B)** and **C)** Representative graphs showing the percentage of viable cells and the fold increase (FI) at each monolayer day, respectively. The percentage of viable cells was calculated as (total number of cells – number of inviable cells)*100/total number of cells. FI was measured as (number of viable cells of Day X) / (number of viable cells plated in Day X-1). Mean values are presented for each day. Error bars were calculated based on standard deviation from two experiments.

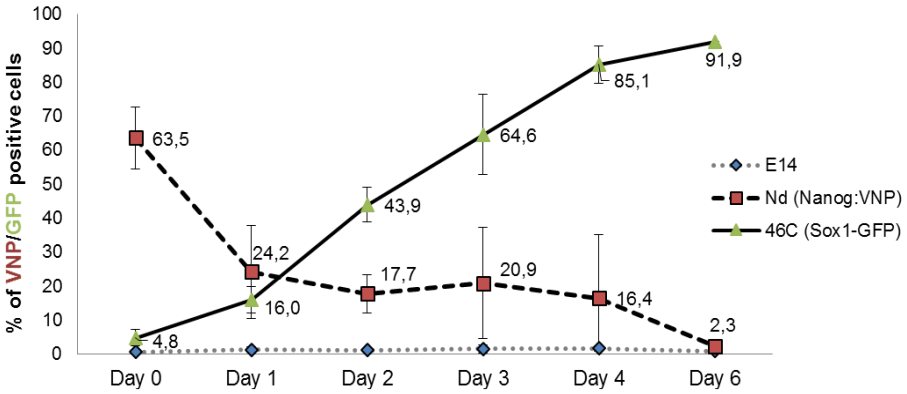



**Figure S3**- Analysis of the reporter proteins expression (Nanog: VNP and Sox1-GFP) by FACS revealed that an increase in the expression of Sox1-GFP reporter as well as a decrease in the expression of Nanog:VNP reporter. Mean values are presented for each day and correspondent cell line. Error bars were calculated based on standard deviation from three experiments.

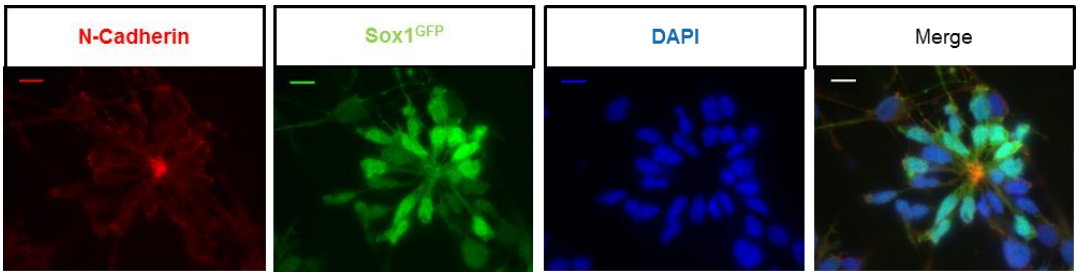



**Figure S4**- Immunofluorescence of rosette-like structures performed at day 6 using antibodies against N-Cadherin (red), marking the apical junctions of these structures, and against GFP (green), marking the neuroepithelial progenitors that express Sox1-GFP reporter. Nuclei were stained with DAPI (blue). (Scale bar = 10µm)

**Table S1:** List of reagents used in the experiments described in this study.

| *Reagent* | Supplier | Cat. Number | Stock | Working stock |
|---|---|---|---|---|
| ***2-βmercaptoethanol*** | Sigma | M-7522 | RT | 0.1M in $H_2O$, 4ºC |
| ***DAPI*** | Sigma | | 1mg/mL in PBS, -20ºC | 1.5µg/mL in PBS, 4ºC |
| ***DMSO*** | Sigma | D-2650 | | RT |
| ***Catalase*** | Sigma | C-3515 | | 4ºC |
| ***ESGRO complete plus*** | Milipore Inc. | SF001-100P | -20ºC | -20ºC |
| ***FBS ES-qualified*** | Invitrogen | 10439-024 | -20ºC | Heat-inactivate, -20ºC |
| ***Gel Red*** | Biotium | 41603-01 | 10000x, RT | 500x, RT |
| ***Gelatin 2%*** | Sigma | G-1393 | 4ºC | 0.1% in PBS, 4ºC |
| ***Glutamine*** | GIBCO | 25030-123 | 200Mm | 100x, -20ºC |
| ***Glucose oxidase*** | Sigma | G2133 | 37mg/mL in 50mM Sodium Acetate, -20ºC | 3.7mg/mL in 50mM Sodium Acetate, -20ºC |
| ***Glycine*** | Sigma | G-7403 | | RT |

| | | | | |
|---|---|---|---|---|
| ***GMEM*** | GIBCO | 21710-025 | 4ºC | 1x, 4ºC |
| ***Laminin*** | Sigma | L-2020 | 400x, -20ºC | 1x, 4ºC |
| ***Mowiol*** | Calbiochem | | | RT |
| ***murine bFGF*** | Prepotech | 100-18B | -20ºC | 4ºC |
| ***Non-essential aminoacids*** | GIBCO | 11140-035 | | 100x, 4ºC |
| ***PDL*** | Sigma | P-7280 | 100x, -20ºC | 1x, 4ºC |
| ***Penicilin-Streptomycin*** | GIBCO | 15140-122 | | 100x, -20ºC |
| ***Propidium iodide*** | Invitrogen | P-3566 | 1mg/mL in PBS, 4ºC | 1ng/mL, 4ºC |
| ***RHB-A*** | Takara Bio Inc | SCS-SF-NB-01 | -20ºC | 4ºC |
| ***rTaq Polymerase*** | GE Healthcare | 27-0798-05 | -20ºC | -20ºC |
| ***SeaKem LE Agarose*** | Lonza | | | |
| ***Sodium pyruvate*** | GIBCO | 11360-039 | 100mM | 100x, -20ºC |
| ***Trypsin*** | GIBCO | 25090-028 | 2.5% (v/v), -20ºC | 0.025% in PBS, 4ºC |

**Table S2:** List of solutions used in the experiments described in this study.

| ***Solution*** | **Components** |
|---|---|
| ***1x TAE buffer*** | 40mM Tris; 1mM EDTA; 0.35%glacial acetic acid |
| ***4% PFA*** | 4% (w/v) paraformaldehyde in PBS |
| ***Antifade buffer*** | 1% (v/v) catalase and 1% (v/v) glucose oxidase in glox buffer |
| ***Blocking solution*** | 10%(w/v) FBS; TBST |
| ***Gelatin 0.1%*** | 2% gelatin, PBS |
| ***Glox buffer*** | 85%(v/v) $H_2O$, 10%(v/v) 20xSSC, 4% (v/v) glucose, 1% (v/v) Tris 1M pH=8, 1% (v/v) Triton |
| ***GMEM 1x*** | 80% (v/v) GMEM; 1% (v/v) Glutamine; 1% (v/v) Pen-Strep; 1% (v/v) Sodium Pyruvate; 1% (v/v) non-essential Aminoacids; 10% (v/v) inactivated FBS; 0.001% (v/v) of 2-mercaptoethanol |
| ***Hybridization buffer*** | 14% (w/v) dextran sulfate in $H_2O$, 10% (v/v) formamide, 10% (v/v) 20x SSC |
| ***Loading buffer*** | 60% (v/v) Glycerol; 10Mm EDTA; 0.2% Orange G |
| ***Mowiol mounting medium*** | 0.1% Mowiol; 33% glycerol; 0.1M Tris, pH 8.5 |
| ***Solution A*** | 10mM Tris-HCl pH 8.3; KCl 100mM; $MgCl_2$ 2.5mM |
| ***Solution B*** | 10mM Tris-HCl pH 8.3; $MgCl_2$ 2.5mM; 1% (v/v) Tween20; 1% (v/v) Triton x100; 120µg/mL proteinase K |
| ***TBST*** | 20mM Tris-HCl pH 8; 150mM NaCl; 0.05% Tween-20 |
| ***Trypsin 0.025%*** | 0.25% Trypsin; PBS |
| ***Trypsin 0.25%*** | 2.5%(v/v) Trypsin; 0.01% (v/v) chicken serum; 0.02% (v/v) 0.5M EDTA; PBS |
| ***Wash buffer for PCR*** | 10mM Tris-HCl pH 8.3; KCl 50mM; $MgCl_2$ 1.5mM |
| ***Wash buffer for sm FISH*** | 10% (v/v) 20x SSC, 10% (v/v) formamide; $H_2O$ |

**Table S3:** List of antibodies used in the experiments described in this study.

| ***Antibody anti-*** | **Dilution** | **Animal** | **Origin** |
|---|---|---|---|
| ***N-cadherin*** | 1:200 | Mouse | BD Transduction Lab #610920 |

C

| | | | |
|---|---|---|---|
| ***GFP*** | 1:1000 | Rabbit | Abcam #ab290 |
| ***Mouse*** | 1:400 | Goat | Molecular Probes #A11032 |
| ***Rabbit*** | 1:400 | Goat | Molecular Probes |

**Table S4:** List of probes used for single molecule RNA FISH. All probes were ordered to Biosearch Technologies and its hybridization temperature is 37ºC

| *Gene* | Probe set |
|---|---|
| ***Car2*** | gtgacaggcagaggtgacag; aaggggaggagaccgtggag; tgattggggcagagcagaag; ctccattggcaatggggaag; tgctgtgtcaatgtccacag; gatatgagcagaggctgtag; gttgacaatgctcttggacg; tcaacgttaaaggagtggcc; attgtcctgagagtcatcaa; atctgtaggagtcactgagg; ccccagtgaaagtgaaactg; gttccagtgaaccaagtgaa; caaaacagccaatccatccg; gaagttagcaaaggccgcac; caggaagaagggagcaagga; tatgtccagtagtccaagtt; acgatccaggtcacacattc; ctcgctgctgacagtaatgg; cattgaagttcagcgtacgg; ctattctttagcggctgagc; cttaaaggacgctttgatct; ttagctacagagaggcggtc; caaatcacccagcctaactg; acaataccagatgcgagtcg; agcacaacggatgagaggta; gtctcatgatgtggacttgt; tttgcctaagttacttcagt; atccattgtgttgtggtatg |
| ***Cdh2*** | gactgcgctgtaaacatctt; tcgctgctttcatactgaac; aagcttctcacagcatacac; tggagttttctggcaagttg; caagctcttgaggaaagggt; tcctgagatggggttgataa; agaggctttgtgactgacag; ctcaggtctgttatcattca;tggaacagacccattccaaa; ttggatcatccgcatcaatg; acatgttgggtgaaggtgtg; ccagtctcattgttgattgt: ctgtggctgtgtttgaaagg; ctctggaggattgtcattga; gacagttaggttggctacaa; ctgcagcaacagtaaggaca; tgctgaatgccttagctaa; tgattttaggatttggggca; tgttgcatatatcgatcggg; attcacggggtctattttca; cggcaatagtagtgatctgc; tacatttggcgattctctgt; atagatttgcagtgttcctg; ctcttgaggtaacacctgag; aagatcaaacgcgaacggcc; tgatggtccagtttctttta; tcatagataccagcttccaa; gattccctgaatctgtgata; aactttcacacgcaggatgg; gaaggatagcgatgatggcg; cgtttcatccataccacaaa; ttaaaagctgcttggcttgg; taagctgtagaggcggggat; agggggctagagataaaggt; tggggacaaacgttgaaggc; cagaaggtagctcacgtaca; gagctaggatacatgctcac; gcggaggactgacaaagttc; tatttgtcttttgtcctggt; tgatggtctgtgtacaccac; cctgatggtgatgtatcgac; tctgcggttctgaaaccaaa; aaggaggtcacaggactcaa; ggacagaaccaggcactgag; acagtggagtttaaaacccc; atctcttctgaggattctcg; tagatgctgttcgtgggtag; aggaatcacttcgtttgtcg; cctggcaaagagcatgaacg; tgatagcttcatgtcggtag; gcaaagagctggggcaaaag; ctcagctccacacagaaata; caccttggctcaagtattgc; ttctcagaaagcagtcagca; tagtttgaatagtccctggc; tgtggagaagaaagcccaga; taaacaagcccttgtcgcag; acgagggaagggacaggaag; aacaattccggtcttcttca; gatcagtgactcgaacagca; aaaaggcttgtttggctcgt; ccatggatcaagaaggcttt; cttcaccatacaacttctct; gttcgtttatctttgtcttt |
| ***Cldn6*** | gctaagatttccaggtcgaa; gatttgcagaccagtagagg; caagcagggtcaggacgatc; cgatgctgttgccgatgaag; acagtgagtcatacaccttg; cacacaggtagtgcacttgg; acgagacttggagttcctat; atgatgccagagatgagcac; caggaccccagaaatgacaa; gtccagcagacaggaatgag; gttgtagaagtcctggatga; tttgagcatcagccaccaag; aatagcaggccatgtaatgt; ccgagaatgtgggacagatg; attcttggtgggatattcgg; taacctttcttgatggctca; acagacagggcaccacatga; aggcaggaaaactggagctg; cagcctcagtagctaaaagt; aaaaacaaaggccaggcagc; ggactgggacaggatatttt;ttccttggtaacttatcacc; ccaaaggttggaagtctcag; taaccttggcgatgggattc; acctttttgtagggagagtac; attgtgtgagcagggaagtg |
| ***Crabp2*** | ccagagctctaggcttttat; ccagagctctaggcttttat; cctttgcagaacagatcctt; ggctaaagatactttgctgt; ttcaactagaacactggacg; gcaaggtggctttctcttta; ttgccagaaaagttaggcat; agttttccgatcggatgatc; tctcctgtttgatctcgact; aagttaatctccgtggttcg; caccaaactcttacagggtc; attggtcagttctcggctcc; ttgtcaggatcagctctcca; tagaccctggtgcacacaac; cgtaggcactcactctcgga; gaagtcgtctcaggcagttc; tttgtgtcctgtagcgggca; agtaaccctcaggagtaagg; gagggttctgagcttctgcc; ttccaggaaggaccctggtg; tagaggccagagagacagtc; aatcacacagactacaagga |
| ***Dnmt3b*** | attcagatgtctgctgtctc; ataatgcactcctcataccc; ggtcactgaagttcccatta; taattcagaaggctggagac; atcatctctgtctccatctc; agccattcccatcatctact; ggtgagctttggcattagaa; ttgctgaagatgatgctcga; acttcttccatgaagtcgac; aagtcaactgatggggtact; tatccataccctcctgatct; atactctgtgctgtctccat; ggtcacctattccaaactct; aggagaagcccttgatcttt; aacttgccatcaccaaacca; tttgtcagcagagatctcag; ttaaagtgctggctgaacag; cagcttattgaaggtagcca; gtacatggccttcctataag; tgcttcttgttgggtttgag; gcaccttcgacttattaacc; aagttcctactgtctgaacg; ttcgacttttgttctcgcgt; aagcagcagagtcattggtt; tatttgtcttgaggcgcttg; ttggtgacttcagaagccat; aaacagcggtcttccagatt; acagggttcttctttccaca; tgtagaagagctctaggaag; tgatagccgtcctcatcata; atctgcagcagcttgtgtta; tgttccaatctttcctgcgt; tagtgaagaagtcttgcagg; aattcttccaggtcaggatc; aattgctgggtacaacttgg; actctaatgggcctcctttt; ccgttgcaattccatcaaac; aactccttgagcaccaagta; gaggcaatgtacttttccac; ttaacagttcccacagcgat; ccggacgtcattgacatatt; ggccccactcttcaatattt; agagatcattgcatgggctt; atataaacctttgcgggcag; aaactcgaagaagagccttc; gggtataattcagcaagtgg; atccagaagaatggacggtt; ttcatggccacaacattctc |

| | |
|---|---|
| ***Fgf5*** | caagcgctgtggatcaggtg; cttcgggagtgagacgcttc; gagtctcccgggttcctagg; aagaaaacgtcgcgctactt; ctatgttccgagccgcttcc; cgaagggctccactggaaac; ccactctgcagtacaggctg; tagatctgcagatggaaacc; agccattgactttgccatcc;cttaacacactggcttcgtg; ttcctacaatcccctgagac; ttgttgctgaaaactcctcg; agtcatccgtaaatttggca; tcttggaatctctccctgaa; gcggacgcataggtattata; cgtaccactctcggcctgtc; gagacgtgttggggtttgac; cttgaacctgggtaggaagt; gtgaaggaaagttccggttg; cacctttggtttcaccggtg; tcagtctgtacttcactggg; gagcatcatccaaagcgaaa; tgacgcctgtatagagagtt; gctgtatccgagtttccttc; aaatgacctgacttccacac; tagttccagtcaaagcgaaa; cccacttctgtttcgactta; ctccctggatcgctacagag; atcctcggtgatctgaagag; cctcaaatcacgaagccttt; gaaaacgcagaatgctaacc; gttccaggctatcaaactgg |
| ***Nanog*** | aaatcagcctatctgaaggc; cagaaagagcaagacaccaa; gaagtcagaaggaagtgagc; actcagtgtctagaaggaaa; ggttttaggcaacaaccaaa; cgagggaagggatttctgaa; cacactcatgtcagtgtgat; cagaactaggcaaactgtgg; ttcccagaattcgatgcttc; aaaaactgcaggcattgatg; agcaagaatagttctcggga; cagagcatctcagtagcaga; gaagaggcaggtcttcagag; tgggactggtagaagaatca; tcaggacttgagagcttttg; cttgttctcctcctcctcag; gagaacacagtccgcatctt; ctgtccttgagtgcacacag; tgaggtacttctgcttctga; gagagttcttgcatctgctg; atagctcaggttcagaatgg; gaaaccaggtcttaacctgc; ttgcacttcatcctttggtt; tcaaccactggtttttctgc; ttctgaatcagaccattgct; gatactccactggtgctgag; ggatagctgcaatggatgct; cagatgcgttcaccagatag; aagttgggttggtccaagtc; gtctggttgttccaagttgg; aaagtcctccccgaagttat; ctgcaactgtacgtaaggct; caaatcactggcagagaagt; tagtggcttccaaattcacc; ctaaaatgcgcatggctttc; ataattccaaggcttgtggg; tggagtcacagagtagttca; agatgttgcgtaagtctcat; gctttgccctgactttaagc; tttggaagaaggaaggaacc; caaatcactggcagagaagt; tagtggcttccaaattcacc; ctaaaatgcgcatggctttc; ataattccaaggcttgtggg; tggagtcacagagtagttca; agatgttgcgtaagtctcat; gctttgccctgactttaagc; tttggaagaaggaaggaacc |
| ***Pou5f1*** | tgagaaggcgaagtctgaag; aggttcgaggatccacccag; tggaggcccttggaagctta; tgagcctggtccgattccag; acatggggagatccccaata; tccctccgcagaactcgtat; aacctgaggtccacagtatg; aacttgggggactaggccca; tcaggctgcaaagtctccac; tgctttccactcgtgctcct; tcagaggaggttccctctga; ttctccaacttcacggcatt; tttcatgtcctgggactcct; aactgttctagctccttctg; tcttctgcttcagcagcttg; tgggtgtaccccaaggtgat; aaagagaacgcccagggtga; tggtctggctgaacaccttt; aaggcctcgaagcgacagat; catgttcttaaggctgagct; ttcctccacccacttctcca; gaaggttctcattgttgtcg; gtctccgatttgcatatctc; tagttcgctttctcttccgg; cacctcacacggttctcaat; tcagaaacatggtctccaga; atctgctgtagggagggctt; aagctgattggcgatgtgag; gaaccacatccttctctagc; cgccggttacagaaccatac; acttgatcttttgcccttct; cttctcgttgggaatactca; ggtgtccctgtagcctcata; agaggaaaggatacagcccc; atagcctggggtgccaaagt; gtgtggtgaagtgggggctt; tcaggaaaagggactgagta; aacagagggaaaggcctcgc; atgggagagcccagagcagt; gctggtgcctcagtttgaat; ttgccttggctcacagcatc; aaagctccaggttctcttgt; ccctcctcagtaaaagaatt; ccacccctgttgtgctttta; agcttctttccccatcccac; ctcctgatcaacagcatcac; aatgatgagtgacagacagg; gtgtgtcccagtctttattt |
| ***Sox2*** | ccgtctccatcatgttatac; tccgggctgttcttctggtt; ataccatgaaggcgttcatg; ttctcctgggccatcttacg; atctccgagttgtgcatctt; tcggacaaaagtttccactc; ttataatccgggtgctcctt; tcatgagcgtcttggttttc; ggaagcgtgtacttatcctt; tagctgtccatgcgctggtt; ttgctccagccgttcatgtg; tcctgcatcatgctgtagct; tgcatcggttgcatctgtgc;tcatggagttgtactgcagg; ttcatgtaggtctgcgagct; agtaggacatgctgtaggtg; ttgaccacagagcccatgga; tgggaggaagaggtaaccac; aggtacatgctgatcatgtc; tgggccatgtgcagtctact; agtgtgccgttaatggccgt; aaaatctctcccccttctcca; cccaattcccttgtatctct; tactctcctctttttgcacc; ctgcggagatttttttcct; tttttccgcagctgtcgttt; aatttggatgggattggtgg; tagtcggcatcacggttttt; gaagtcccaagatctctcat; ctgtacaaaaatagtccccc; tatacatggtccgattcccc; gcgtagttttttcctccag; cctaacgtaccactagaact; aagacttttgcgaactccct; ccggagtctagctctaaata; ctgtacaaaagttgcttgca; gattgccatgtttatctcga; caagaacccttttcctcgaaa; aagctgcagaatcaaaaccc; ccttgtttgtaacggtccta; ccagtacttgctctcatgtt; aacaagaccacgaaaacggt; acaatctagaacgtttgcct; gatatcaacctgcatggaca; gggtaggattgaacaaaagc; cggaaaataaaaggggggaa; ccaataacagagccgaatct; tatacatggattctcggcag |
| ***Sox3*** | ttctctcgagctggtcgcat; cgggcttctctcacctgatg; acaccatgaacgcgttcatg; ggttctccagggccatcttg; atctcggagttgtgcatctt; catcggtcagcagtttccag; cttggcctcgtcgatgaacg; gtacttgtagtccgggtact; gagcagcgtcttggtcttgc; gcagcgagtacttgtccttc; ttcacgtgcgtgtacgtgtc; ctcctgcacgagcgagtagg; atgtcgtagcggtgcatctg; gagagctgggctccgacttc; ggtacatgctgatcatgtcg; accgttccattgaccgcagt; gagcaaagctaaacagcaag; catcttcggtacaaggcaac; gacagttacggccaaacttt; ggacttctcgcttttgtaca; gctctagcaagtcccatttc; gaacctaggaatccgggaag; gacattttcaactgcaacag; gggcaacctcactcagttct; tggaggcattgcagttcttg; aacattggctttagctgtcg; aactcaacagcctaaacgcg; agcaaatagatcactgcaga; gaacgaaatgcgtacacgaa; actttgaaaaaacctggaac |
| ***T*** | gtactgcaggctcttccctg; cgctgagcaggtggtccact; ctcgcagttcgcgttcggtg; ttagctccttgaagcgcagc; atcctcctgccgttcttggt; ggcctgacacatttaccttc; gcaagaaagagtacatggca; ttgtcagccgtcacgaagtc; cctgaggctctggtttgccc; atccagtgggccccaaaatt; tgactttgctgaaagacaca; ccattgagcttgttggtgag; aggttcatacttatgcaagg; ccaactctcacgatgtgaat; aagcagtggctggtgatcat; catcaaggaaggctttagca; cctccattacatctttgtgg; cgagcctccaaactgagggt; gtacctctcacagccgtgtg; gatgaccggtggttccttag; gagctgttccgatgagcata; gatgaattgtccgcataggt; ctgcagcatggacagacaag; cgaggctagaccagttatca; ctcgaaagaactgagctccc; ctgagaccgtgtgcgtcagt; tcatacatcggagaaccaga; atgtctgtgactgtagcagc; cgtgtcatactggctgtcag; caggtgtccacgaggctatg; gttcaattcacatagatggg; ttacaagtctcagcacatgg; tagaagatccagttgacacc; caagagcctgccactttgag; aggacatgtattattgctct; acaagaggctgtagaacatg; aggacacatatcactactgt; cacaaaaaactggaccacag; aggactctgattaactgcca; |

| | |
|---|---|
| | aacacaagatataggaccct; ggacgaattccaggatttca; tcactaaagtaggacagggg; acattagaggtgtaccttgt; tactctaaggcaacaaggga; ggctatgctttttctgtcc; cagttgacggttcagttaca; aagctacgtgtgcttaaagt; aagtatccatacagttgact |

**Table S5:** List of mRNA half-lives for the genes analysed in this work.

| | Gene | mRNA half-live (hours) according to Sharova et al. [60] data set |
|---|---|---|
| **Pluripotency genes** | *Oct4* | 7.4 |
| | *Sox2* | 1 |
| | *Nanog* | 5.2 |
| **Stochastic genes** | *Car2* | 5.6 |
| | *Cldn6* | 5.3 |
| **Lineage-affiliated genes** | *Fgf5* | 2.8 |
| | *Dnmt3b* | 5.4 |
| | *Cdh2* | 6.3 |
| | *Sox3* | 4.5 |
| | *Crabp2* | 7.2 |
| | *T* | 2.5 |

**Table S6:** Threshold values defined for each gene analysed in this work.

| | Gene | Thresholds (mRNA molecules per cell) |
|---|---|---|
| **Pluripotency genes** | *Oct4* | **150** |
| | *Sox2* | **100** |
| | *Nanog* | **50** |
| **Stochastic genes** | *Car2* | **100** |
| | *Cldn6* | **30** |
| **Lineage-affiliated genes** | *Fgf5* | **25** |
| | *Dnmt3b* | **50** |
| | *Cdh2* | **25** |
| | *Sox3* | **50** |
| | *Crabp2* | **80** |
| | *T* | **5** |

G